Applied Statistics and Econometrics
Lecture 5

Giuseppe Ragusa

Luiss University

gragusa@luiss.it
http://gragusa.org/

February 27, 2017

Luiss University

Linear Regression Model: General

Although we have been discussing about the testscore and str, the above model is more general, so we will find useful to introduce more general notation:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

is the **linear regression model with a single regressor.**

Linear Regression Model: General

Although we have been discussing about the testscore and str, the above model is more general, so we will find useful to introduce more general notation:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

is the **linear regression model with a single regressor.**

- Y_i is the **dependent variable**
- X_i is the **independent variable** or **regressor**

Linear Regression Model: General

Although we have been discussing about the testscore and str, the above model is more general, so we will find useful to introduce more general notation:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

is the **linear regression model with a single regressor.**

- Y_i is the **dependent variable**
- X_i is the **independent variable** or **regressor**
- $\beta_0 + \beta_1 X_i$ is the **population regression line**

Linear Regression Model: General

Although we have been discussing about the testscore and str, the above model is more general, so we will find useful to introduce more general notation:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

is the **linear regression model with a single regressor.**

- Y_i is the **dependent variable**
- X_i is the **independent variable** or **regressor**
- $\beta_0 + \beta_1 X_i$ is the **population regression line**
- u_i is the **error term** incorporating all the factors responsible for the difference between Y_i and $X_i\beta_1$

Regression functions and conditional expectations

Conditional Expectation

If $E[u_i|X_i] = 0$ then

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_i$$

$\beta_0 + \beta_1 X_i$ is the conditional expectation of Y given X.

Regression functions and conditional expectations

Conditional Expectation

If $E[u_i|X_i] = 0$ then

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_i$$

$\beta_0 + \beta_1 X_i$ is the conditional expectation of Y given X.

Caveats

$E[u_i|X_i] = 0$ is a **big** assumption and we will question it in a bit. However, for now, we all assume it holds.

Example: Class Size

The linear model for testscore and str:

$$testscore_i = \beta_0 + \beta_1 str + u_i$$

Example: Class Size

The linear model for testscore and str:

$$testscore_i = \beta_0 + \beta_1 str + u_i$$

implies that, e.g.,

$$E[testscore_i | str = 20] = \beta_0 + \beta_1 \times 20$$

Example: Class Size

The linear model for testscore and str:

$$testscore_i = \beta_0 + \beta_1 str + u_i$$

implies that, e.g.,

$$E[testscore_i | str = 20] = \beta_0 + \beta_1 \times 20$$

and

$$E[testscore_i | str = 21] = \beta_0 + \beta_1 \times 21.$$

Interpretation of the intercept β_0

Given the linear model for (Y_i, X_i), we have that

$$E[Y_i|X_i = 0] = \beta_0 + \beta_1 \times 0 = \beta_0.$$

Thus, the intercept is the expected value of Y_i when $X_i = 0$.

Interpretation of the intercept β_0

Given the linear model for (Y_i, X_i), we have that

$$E[Y_i|X_i = 0] = \beta_0 + \beta_1 \times 0 = \beta_0.$$

Thus, the intercept is the expected value of Y_i when $X_i = 0$.

Remarks

- It is common for β_0 not to have a meaningful economic interpretation. Take, e.g.,

  $$testscore_i = \beta_0 + \beta_1 str_i + u_i.$$

  Then β_0 is the expected value of *testscore* in a district with $str = 0$, that is a district with no students!

Interpretation of the slope (β_1)

β_1 tells us what is the change of the cond. expectation of Y due to a change of 1 unit of X.

Derivation

The expected value of Y when X_i is set to some value x is

$$E[Y_i|X_i = x] = \beta_0 + \beta_1 \times x.$$

Incrementing the value of X_i of 1 unit, i.e. $X_i = x + 1$, we have

$$E[Y_i|X_i = x + 1] = \beta_0 + \beta_1 \times (x + 1).$$

Thus

$$\begin{aligned} E[Y_i|X_i = x + 1] - E[Y_i|X_i = x] &= (\beta_0 + \beta_1 \times (x + 1)) - (\beta_0 + \beta_1 \times x) \\ &= \beta_1 \end{aligned}$$

Interpretation of the slope (β_1)

If the change in X_i is different from 1 unit, then β_1 needs to be multiplied by the number of unites X_i is changed:

$$E[Y_i|X_i = x] = \beta_0 + \beta_1 \times x.$$

Interpretation of the slope (β_1)

If the change in X_i is different from 1 unit, then β_1 needs to be multiplied by the number of unites X_i is changed:

$$E[Y_i|X_i = x] = \beta_0 + \beta_1 \times x.$$

Incrementing the value of X_i of 1 unit, i.e. $X_i = x + 1$, we have

$$E[Y_i|X_i = x + 5] = \beta_0 + \beta_1 \times (x + 5).$$

Interpretation of the slope (β_1)

If the change in X_i is different from 1 unit, then β_1 needs to be multiplied by the number of unites X_i is changed:

$$E[Y_i|X_i = x] = \beta_0 + \beta_1 \times x.$$

Incrementing the value of X_i of 1 unit, i.e. $X_i = x + 1$, we have

$$E[Y_i|X_i = x + 5] = \beta_0 + \beta_1 \times (x + 5).$$

Thus,

$$\begin{aligned} E[Y_i|X_i = x + 5] - E[Y_i|X_i = x] &= (\beta_0 + \beta_1 \times (x + 5)) - (\beta_0 + \beta_1 \times x) \\ &= \beta_1 \times 5 \end{aligned}$$

Example: testscore vs str

The linear model

$$testscore_i = \beta_0 + \beta_1 str + u_i$$

$$E[testscore_i | str_i = 20] = \beta_0 + \beta_1 \times 20.$$

Incrementing the value of *stri* of 1 unit, i.e. $stri = x + 1$, we have

$$E[testscore_i | str_i = 21] = \beta_0 + \beta_1 \times 21.$$

Thus, the effect of **increasing** str by 1 unit is

$$E[testscore_i | str_i = 21] - E[testscore_i | str_i = 20]$$
$$= (\beta_0 + \beta_1 \times 21) - (\beta_0 + \beta_1 \times 20) = \beta_1$$

The effect of **decreasing** str by 1 unit is

$$E[testscore_i | str_i = 20] - E[testscore_i | str_i = 21] = -\beta_1$$

Estimation of β_0 and β_1

We already know how to fit a line through points....

$$
\begin{aligned}
\{\hat{\beta}_0, \hat{\beta}_1\} &= \underset{\beta_0,\beta_1}{\operatorname{argmin}}\, u_1^2 + u_2^2 + \ldots + u_n^2 \\
&= \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} u_i^2 \\
&= \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - \beta_0 - X_i\beta_1)^2
\end{aligned}
$$

Which gives:

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \bar{X}\hat{\beta}_1 \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\widehat{\mathrm{cov}}(Y, X)}{\widehat{\mathrm{var}}(X)}
\end{aligned}
$$

The OLS Estimator, Predicted Values, and Residuals

The OLS estimators of the slope β_1 and the intercept β_0 are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.7}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.8}$$

The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1, \ldots, n \tag{4.9}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{4.10}$$

The estimated intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$), and residual ($\hat{u}_i$) are computed from a sample of n observations of X_i and Y_i, $i = 1, \ldots, n$. These are estimates of the unknown true population intercept (β_0), slope (β_1), and error term (u_i).

Implementing (Ordinary) Least Squares

Given data, $\hat{\beta}_0$ and β_1 are obtained by using a computer program (In our case, `R`). The output is written as:

$$\widehat{testscore} = 698.9329 - 2.2798 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$;

Implementing (Ordinary) Least Squares

Given data, $\hat{\beta}_0$ and β_1 are obtained by using a computer program (In our case, `R`). The output is written as:

$$\widehat{testscore} = 698.9329 - 2.2798 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$;
- Estimated intercept: $\hat{\beta}_0 = 698.9329$;

Implementing (Ordinary) Least Squares

Given data, $\hat{\beta}_0$ and β_1 are obtained by using a computer program (In our case, `R`). The output is written as:

$$\widehat{testscore} = 698.9329 - 2.2798 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$;
- Estimated intercept: $\hat{\beta}_0 = 698.9329$;
- $\widehat{testscore}$ denotes the estimated regression line.

R: testscore vs str

```
## Run the linear model function
lm(testscore ~ str, data = CASchools)
```

```
##
## Call:
## lm(formula = testscore ~ str, data = CASchools)
##
## Coefficients:
## (Intercept)          str
##      698.93        -2.28
```

More about R and the linear model later.

OLS Interpretation

- Remember that β_1 is the change of the expectation of *testscore* due to a unit change of *str*
- We do not know β_1, instead we use $\hat{\beta}_1$
- Thus, an **increase** of str of 1 leads to an **estimate** decrease of the expected value of *testscore* of -2.28 points.

Measures of fit

How well the linear regression describes the data?

- The R^2 and the SER measure ho well the OLS regression line fits the data.

R^2

The regression R^2 is the fraction of the sample variance of Y_i "explained" by X_i.

Definition of R^2

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{\widehat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- $R^2 = 0$ means $ESS = 0$;
- $R^2 = 1$ means $ESS = TSS$;
- $0 \leqslant R^2 \leqslant 1$;
- For regression with a single X, $R^2 =$ the square of the correlation coefficient between X and Y

R^2 with R

```
lm_1 <- lm(testscore ~ str, data = CASchools)
Y <- CASchools$testscore
Yhat <- predict(lm_1)
Ybar <- mean(Y)
Yhatbar <- mean(Yhat)
ESS <- sum((Yhat - Yhatbar)^2)
TSS <- sum((Y - Ybar)^2)
R2 <- ESS/TSS
R2

## [1] 0.0512
```

Standard error of the regression *SER*

The Standard Error of the Regression (SER) measures the spread of the distribution of $\hat{u}$. The SER is (almost) the sample standard deviation of the OLS residuals:

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}.$$

The SER:

- has the units of u, which are the units of Y;
- measures the average "size" of the OLS residual (the average "mistake" made by the OLS regression line)
- The root mean squared error (RMSE) is closely related to the SER:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}.$$

SER with R

```
lm_1 <- lm(testscore ~ str, data = CASchools)
u <- resid(lm_1)
SER <- sqrt(sum(u^2)/(length(u) - 2))
SER

## [1] 18.6
```

summary

```
lm_1 <- lm(testscore ~ str, data = CASchools)
summary(lm_1)
```

```
##
## Call:
## lm(formula = testscore ~ str, data = CASchools)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -47.73 -14.25   0.48  12.82  48.54
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   698.93       9.47   73.82  < 2e-16 ***
## str            -2.28       0.48   -4.75  2.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.6 on 418 degrees of freedom
## Multiple R-squared:  0.0512,Adjusted R-squared:  0.049
## F-statistic: 22.6 on 1 and 418 DF,  p-value: 2.78e-06
```

The Least Squares Assumptions (SW Section 4.4)

- What, in a precise sense, are the properties of the OLS estimator? We would like it to be unbiased, and to have a small variance. Does it?

The Least Squares Assumptions (SW Section 4.4)

- What, in a precise sense, are the properties of the OLS estimator? We would like it to be unbiased, and to have a small variance. Does it?
- Under what conditions is it an unbiased estimator of the true population parameters?

The Least Squares Assumptions (SW Section 4.4)

- What, in a precise sense, are the properties of the OLS estimator? We would like it to be unbiased, and to have a small variance. Does it?
- Under what conditions is it an unbiased estimator of the true population parameters?
- To answer these questions, we need to make some assumptions about how Y and X are related to each other, and about how they are collected (the sampling scheme)

The Least Squares Assumptions (SW Section 4.4)

- What, in a precise sense, are the properties of the OLS estimator? We would like it to be unbiased, and to have a small variance. Does it?
- Under what conditions is it an unbiased estimator of the true population parameters?
- To answer these questions, we need to make some assumptions about how Y and X are related to each other, and about how they are collected (the sampling scheme)
- These assumptions— there are three —are known as the Least Squares Assumptions.

The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n$$

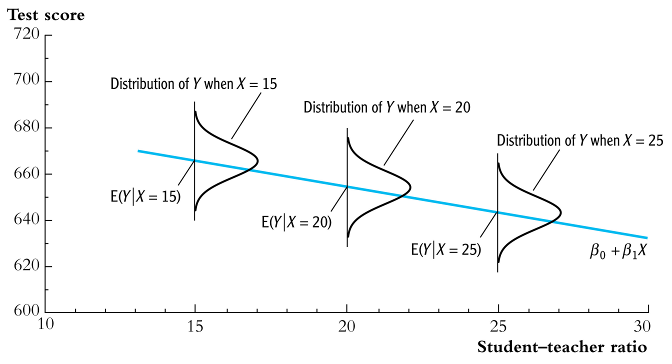
- The conditional distribution of u given X has mean zero, that is,

$$E(u_i | X_i = x) = 0$$

- (Y_i, X_i) are iid
- Large outliers in Y and X are rare

Least squares assumption 1: $E(u_i|X_i = x) = 0$.

For any given value of X, the mean of u is zero:



Least squares assumption 1, ctd

A benchmark for thinking about this assumption is to consider an ideal randomized controlled experiment:

- X is randomly assigned to people (students randomly assigned to different size classes; patients randomly assigned to medical treatments). Randomization is done by computer – using no information about the individual.

Least squares assumption 1, ctd

A benchmark for thinking about this assumption is to consider an ideal randomized controlled experiment:

- X is randomly assigned to people (students randomly assigned to different size classes; patients randomly assigned to medical treatments). Randomization is done by computer – using no information about the individual.
- Because X is assigned randomly, all other individual characteristics – the things that make up u – are independently distributed of X

Least squares assumption 1, ctd

A benchmark for thinking about this assumption is to consider an ideal randomized controlled experiment:

- X is randomly assigned to people (students randomly assigned to different size classes; patients randomly assigned to medical treatments). Randomization is done by computer – using no information about the individual.
- Because X is assigned randomly, all other individual characteristics – the things that make up u – are independently distributed of X
- Thus, in an ideal randomized controlled experiment, $E(u_i|X_i = x) = 0$ holds.

Least squares assumption 1, ctd

A benchmark for thinking about this assumption is to consider an ideal randomized controlled experiment:

- X is randomly assigned to people (students randomly assigned to different size classes; patients randomly assigned to medical treatments). Randomization is done by computer – using no information about the individual.
- Because X is assigned randomly, all other individual characteristics – the things that make up u – are independently distributed of X
- Thus, in an ideal randomized controlled experiment, $E(u_i|X_i = x) = 0$ holds.
- In actual experiments, or with observational data, we will need to think hard about whether $E(u_i|X_i = x) = 0$ holds.

Note

Fact

If $E(u|X = x) = 0$, then $\operatorname{cov}(u, X) = 0$. The converse is not true.

Thus, checking whether $E(u|X = x) = 0$ can be done by checking whether $\operatorname{cov}(u, X) = 0$.

In particular, Assumption 1 will be violated if the other factors are correlated with X. (Again, $\operatorname{cov}(u, X)$ maybe 0, but $E(u|X) \neq 0$).

Least squares assumption 2: (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d.

- This arises automatically if the entity (individual, district) is sampled by simple random sampling: the entity is selected then, for that entity, X and Y are observed (recorded).

Least squares assumption 2: (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d.

- This arises automatically if the entity (individual, district) is sampled by simple random sampling: the entity is selected then, for that entity, X and Y are observed (recorded).
- The main place we will encounter non-i.i.d. sampling is when data are recorded over time ("time series data") – this will introduce some extra complications.

Least squares assumption 3: Large outliers are rare

$$E(Y^4) < \infty, \quad E(X^4) < \infty$$

- A large outlier is an extreme value of X or Y

Least squares assumption 3: Large outliers are rare

$$E(Y^4) < \infty, \quad E(X^4) < \infty$$

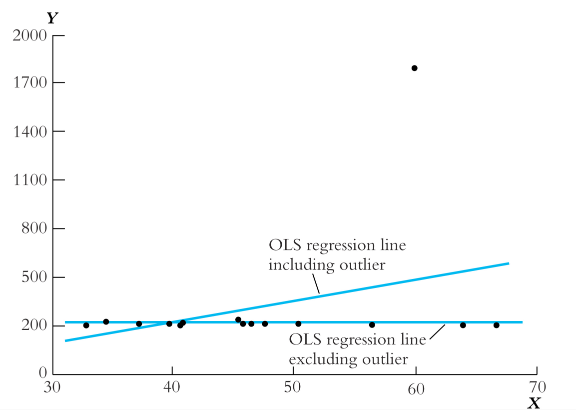
- A large outlier is an extreme value of X or Y
- On a technical level, if X and Y are bounded, then they have finite fourth moments. (Standardized test scores automatically satisfy this; str, family income, etc. satisfy this too).

Least squares assumption 3: Large outliers are rare

$$E(Y^4) < \infty, \quad E(X^4) < \infty$$

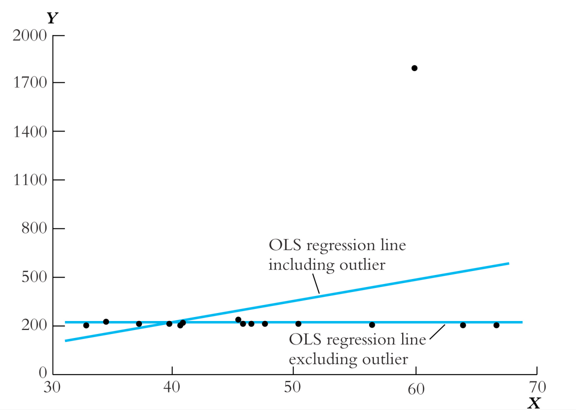
- A large outlier is an extreme value of X or Y
- On a technical level, if X and Y are bounded, then they have finite fourth moments. (Standardized test scores automatically satisfy this; str, family income, etc. satisfy this too).
- However, the substance of this assumption is that a large outlier can strongly influence the results

Outliers



- Is the lone point an outlier in X or Y?

Outliers



- Is the lone point an outlier in X or Y?
- In practice, outliers often are data glitches (coding/recording problems) – so check your data for outliers!

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;
- use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$;

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;
- use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$;
- construct a confidence interval for β_1;

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;
- use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$;
- construct a confidence interval for β_1;
- All these require figuring out the sampling distribution of the OLS estimator. Two steps to get there...

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;
- use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$;
- construct a confidence interval for β_1;
- All these require figuring out the sampling distribution of the OLS estimator. Two steps to get there...
  - Probability framework for linear regression;

The Sampling Distribution of the OLS Estimator (SW Section 4.5)

The OLS estimator is computed from a sample of data; a different sample gives a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$.

We want to:

- quantify the sampling uncertainty associated with $\hat{\beta}_1$;
- use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$;
- construct a confidence interval for β_1;
- All these require figuring out the sampling distribution of the OLS estimator. Two steps to get there...
    - Probability framework for linear regression;
    - Distribution of the OLS estimator

Probability Framework for Linear Regression

Population

The group of interest (ex: all possible school districts)

Random variables

Y, X Ex: (Test Score, str)

Joint distribution of (Y, X)

The population regression function is linear $E(u_i|X_i) = 0$ (1st Least Squares Assumption)

X, Y have finite fourth moments (3rd L.S.A.)

Data Collection by simple random sampling

(X_i, Y_i), $i = 1, \ldots, n$ are i.i.d. (2nd L.S.A.)

The Sampling Distribution of $\hat{\beta}_1$

Like $\bar{Y}$, $\hat{\beta}_1$ has a sampling distribution.

- What is $E(\hat{\beta}_1)$?
- What is $\mathrm{var}(\hat{\beta}_1)$?
- What is the distribution of $\hat{\beta}_1$ in small samples?
- What is the distribution of $\hat{\beta}_1$ in large samples?

Mean and variance of the sampling distribution of $\hat{\beta}_1$

Preliminary algebra

Given $Y_i = \beta_0 + \beta_1 X_i + u_i$ and taking means on both sides, noting that

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$$

we can express the model in mean deviation

$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u}). \tag{1}$$

Substituting (1) into the expression for $\hat{\beta}_1$, we get

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Mean and variance of the sampling distribution of $\hat{\beta}_1$, ctd.

We have that

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u}) &= \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\sum_{i=1}^{n}(X_i - \bar{X})\right]\bar{u} \\
&= \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\sum_{i=1}^{n}X_i - n\bar{X})\right]\bar{u} \\
&= \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[n\bar{X} - n\bar{X})\right]\bar{u} \\
&= \sum_{i=1}^{n}(X_i - \bar{X})u_i
\end{aligned}$$

Mean and variance of the sampling distribution of $\hat{\beta}_1$, ctd.

Expected value of $\hat{\beta}_1$

$$
\begin{aligned}
E\left[\hat{\beta}_1 - \beta_1\right] &= E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] \\
&= E\left\{E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\middle| X_1, \ldots, X_n\right]\right\} \\
&= E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})E(u_i|X_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] \\
&= 0
\end{aligned}
$$

Mean and variance of the sampling distribution of $\hat{\beta}_1$

- $E[\hat{\beta}_1 - \beta_1] = 0 \implies E[\hat{\beta}_1] = \beta_1$
- LSA $\#1$ implies that $\hat{\beta}_1$ is unbiased;
- For details see App. 4.3

Mean and variance of the sampling distribution of $\hat{\beta}_1$

Next, calculate $\mathrm{var}(\hat{\beta}_1)$.

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

If $n \to \infty$, $\frac{(n-1)}{n}\hat{\sigma}_X^2 \xrightarrow{p} \sigma_X^2$, and $(X_i - \bar{X})u_i \xrightarrow{p} (X_i - \mu_X)u_i$. Thus,

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n}\sum_{i=1}^{n}\nu_i}{\sigma_X^2},$$

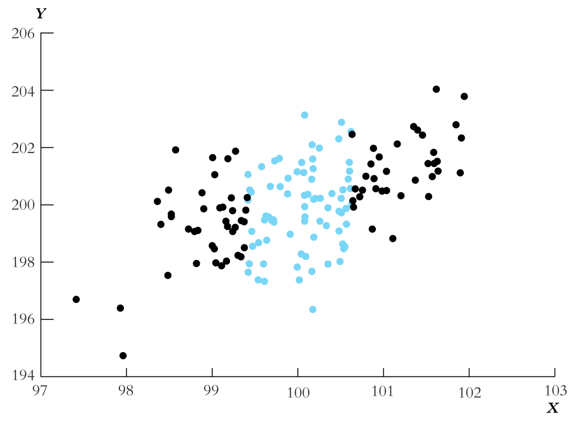
where $\nu_i = (X_i - \mu_X)u_i$.

Mean and variance of the sampling distribution of $\hat{\beta}_1$

$$\mathrm{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\mathrm{var}(\nu_i)}{\sigma_X^4} = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4}$$

- The variance of $\hat{\beta}_1$ is inversely proportional to n— just like $\mathrm{var}(\bar{Y})$.
- The larger the variance of X, the smaller the variance of $\hat{\beta}_1$
    - **The intuition:** If there is more variation in X, then there is more information in the data that you can use to fit the regression line.

The larger the variance of X, the smaller the variance of $\hat{\beta}_1$



There are the same number of black and blue dots – using which would you get a more accurate regression line?

Large sample distribution of $\hat{\beta}_1$

The exact sampling distribution is complicated – it depends on the population distribution of (Y, X) – but when n is large we get some simple (and good) approximation:

$$\hat{\beta}_1 \xrightarrow{d} N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right).$$

where

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4}.$$

Parallel btw the asymptotic distribution of β_1 and $\bar{Y}$

$\hat{\beta}_1$

- $E[\hat{\beta}_1] = \beta_1$
- $\hat{\beta}_1 \xrightarrow{p} \beta_1$
- $\hat{\beta}_1 \xrightarrow{d} N(\beta_1, \sigma^2_{\hat{\beta}_1})$
- $\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma^4_X}$

$\bar{Y}$

- $E[\bar{Y}] = \mu_Y$
- $\bar{Y} \xrightarrow{p} \mu_Y$
- $\bar{Y} \xrightarrow{d} N(\mu_Y, \sigma^2_{\bar{Y}})$
- $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n}$

Summary of the sampling distribution of $\hat{\beta}_1$

If the three LS assumptions hold, then

- The exact (finite sample) sampling distribution of $\hat{\beta}_1$ has:

$$E(\hat{\beta}_1) = \beta_1 \quad \text{(that is, } \hat{\beta}_1 \text{ is unbiased)}$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4} \propto \frac{1}{n}$$

Summary of the sampling distribution of $\hat{\beta}_1$

If the three LS assumptions hold, then

- The exact (finite sample) sampling distribution of $\hat{\beta}_1$ has:

$$E(\hat{\beta}_1) = \beta_1 \quad \text{(that is, } \hat{\beta}_1 \text{ is unbiased)}$$

$$\operatorname{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\operatorname{var}((X_i - \mu_X)u_i)}{\sigma_X^4} \propto \frac{1}{n}$$

- Other than its mean and variance, the exact distribution of is complicated and depends on the distribution of (X, u);

Summary of the sampling distribution of $\hat{\beta}_1$

If the three LS assumptions hold, then

- The exact (finite sample) sampling distribution of $\hat{\beta}_1$ has:

$$E(\hat{\beta}_1) = \beta_1 \quad \text{(that is, } \hat{\beta}_1 \text{ is unbiased)}$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4} \propto \frac{1}{n}$$

- Other than its mean and variance, the exact distribution of is complicated and depends on the distribution of (X, u);
- $\hat{\beta}_1 \xrightarrow{p} \beta_1$, (that is, $\hat{\beta}_1$ is consistent)

Summary of the sampling distribution of $\hat{\beta}_1$

If the three LS assumptions hold, then

- The exact (finite sample) sampling distribution of $\hat{\beta}_1$ has:

$$E(\hat{\beta}_1) = \beta_1 \quad \text{(that is, } \hat{\beta}_1 \text{ is unbiased)}$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4} \propto \frac{1}{n}$$

- Other than its mean and variance, the exact distribution of is complicated and depends on the distribution of (X, u);
- $\hat{\beta}_1 \xrightarrow{p} \beta_1$, (that is, $\hat{\beta}_1$ is consistent)
- When n is large

$$\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{\mathrm{var}(\hat{\beta}_1)}} \sim N(0, 1)$$

Summary of the sampling distribution of $\hat{\beta}_1$

If the three LS assumptions hold, then

- The exact (finite sample) sampling distribution of $\hat{\beta}_1$ has:

$$E(\hat{\beta}_1) = \beta_1 \quad \text{(that is, } \hat{\beta}_1 \text{ is unbiased)}$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\mathrm{var}((X_i - \mu_X)u_i)}{\sigma_X^4} \propto \frac{1}{n}$$

- Other than its mean and variance, the exact distribution of is complicated and depends on the distribution of (X, u);
- $\hat{\beta}_1 \xrightarrow{p} \beta_1$, (that is, $\hat{\beta}_1$ is consistent)
- When n is large

$$\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{\mathrm{var}(\hat{\beta}_1)}} \sim N(0, 1)$$

- This parallels the sampling distribution on $\bar{Y}$.

Large sample distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

Large-Sample Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

If the least squares assumptions in Key Concept 4.3 hold, then in large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ have a jointly normal sampling distribution. The large-sample normal distribution of $\hat{\beta}_1$ is $N(\beta_1, \sigma^2_{\hat{\beta}_1})$, where the variance of this distribution, $\sigma^2_{\hat{\beta}_1}$, is

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\operatorname{var}[(X_i - \mu_X)u_i]}{[\operatorname{var}(X_i)]^2}. \tag{4.21}$$

The large-sample normal distribution of $\hat{\beta}_0$ is $N(\beta_0, \sigma^2_{\hat{\beta}_0})$, where

$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{\operatorname{var}(H_i u_i)}{[E(H_i^2)]^2}, \text{ where } H_i = 1 - \left(\frac{\mu_X}{E(X_i^2)}\right)X_i. \tag{4.22}$$

We are now ready to turn to hypothesis tests & confidence intervals...

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;
- Also, we will cover some loose ends about regression:

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;
- Also, we will cover some loose ends about regression:
    - Regression when X is binary (0/1);

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;
- Also, we will cover some loose ends about regression:
  - Regression when X is binary (0/1);
  - Heteroskedasticity and homoskedasticity (this is new)

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;
- Also, we will cover some loose ends about regression:
    - Regression when X is binary (0/1);
    - Heteroskedasticity and homoskedasticity (this is new)
    - Efficiency of the OLS estimator (also new);

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals (SW Chapter 5)

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1;
- Also, we will cover some loose ends about regression:
  - Regression when X is binary (0/1);
  - Heteroskedasticity and homoskedasticity (this is new)
  - Efficiency of the OLS estimator (also new);
  - Use of the t-statistic in hypothesis testing (new but not surprising)

Hypothesis Testing and the Standard Error of $\hat{\beta}_1$

Objective

The objective is to test a hypothesis, like $\beta_1 = 0$, using data – to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.

General Setup

- Null hypothesis and two-sided alternative:

$$H0 : \beta_1 = \beta_{1,0} \quad vs. \quad H_1 : \beta_1 \neq \beta_{1,0}$$

  where $\beta_{1,0}$ is the hypothesized value under the null.
- Null hypothesis and one-sided alternative:

$$H0 : \beta_1 = \beta_{1,0} \quad vs. \quad H_1 : \beta_1 < (>) \beta_{1,0}$$

General approach to testing

Construct t-statistic, and compute p-value (or compare to N(0,1) critical value)

$$t = \frac{\text{Estimator - Hypothesized value}}{\text{Standard Error of the estimator}}$$

where the SE of the estimator is the square root of an estimator of the variance of the estimator.

General approach to testing

Construct t-statistic, and compute p-value (or compare to N(0,1) critical value)

$$t = \frac{\text{Estimator - Hypothesized value}}{\text{Standard Error of the estimator}}$$

where the SE of the estimator is the square root of an estimator of the variance of the estimator.

- For testing the mean of Y

$$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}}$$

General approach to testing

Construct t-statistic, and compute p-value (or compare to N(0,1) critical value)

$$t = \frac{\text{Estimator - Hypothesized value}}{\text{Standard Error of the estimator}}$$

where the SE of the estimator is the square root of an estimator of the variance of the estimator.

- For testing the mean of Y

$$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}}$$

- For testing β_1,

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

where $SE(\hat{\beta}_1)$ = the square root of an estimator of the variance of the sampling distribution of $\hat{\beta}_1$.

Formula for $SE(\hat{\beta}_1)$

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\mathrm{var}(\hat{\beta}_1) = \frac{\mathrm{var}\left[(X_i - \mu_X)u_i\right]}{n\sigma_x^4} = \frac{\sigma_\nu^2}{n\sigma_X^4}, \text{ where } \nu_i = (X_i - \mu_X)u_i$$

Formula for $SE(\hat{\beta}_1)$

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\mathrm{var}(\hat{\beta}_1) = \frac{\mathrm{var}\left[(X_i - \mu_X)u_i\right]}{n\sigma_x^4} = \frac{\sigma_\nu^2}{n\sigma_X^4}, \text{ where } \nu_i = (X_i - \mu_X)u_i$$

The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_ν^2 and σ_X^4 by estimators constructed from the data:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_\nu^2}{(\text{estimator of } \sigma_X^4)}$$

Formula for $SE(\hat{\beta}_1)$

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\mathrm{var}(\hat{\beta}_1) = \frac{\mathrm{var}\left[(X_i - \mu_X)u_i\right]}{n\sigma_x^4} = \frac{\sigma_\nu^2}{n\sigma_X^4}, \text{ where } \nu_i = (X_i - \mu_X)u_i$$

The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_ν^2 and σ_X^4 by estimators constructed from the data:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_\nu^2}{(\text{estimator of } \sigma_X^4)} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^n \hat{\nu}_i}{\frac{1}{n}\sum_i^n (X_i - \bar{X})^4}$$

Formula for $SE(\hat{\beta}_1)$

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}\left[(X_i - \mu_X)u_i\right]}{n\sigma_x^4} = \frac{\sigma_\nu^2}{n\sigma_X^4}, \text{ where } \nu_i = (X_i - \mu_X)u_i$$

The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_ν^2 and σ_X^4 by estimators constructed from the data:

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \times \frac{\text{estimator of } \sigma_\nu^2}{(\text{estimator of } \sigma_X^4)} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^n \hat{\nu}_i}{\frac{1}{n}\sum_i^n (X_i - \bar{X})^4}$$

where $\hat{\nu}_i = (X_i - \bar{X})\hat{u}_i$.

Formula for $SE(\hat{\beta}_1)$

Standard Error

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}} = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{\nu}_i}{\frac{1}{n}\sum_i^n (X_i - \bar{X})^4}}$$

Remarks

OK, this is a bit nasty, but:

- It is less complicated than it seems. The numerator estimates var(v), the denominator estimates var(X).

Formula for $SE(\hat{\beta}_1)$

Standard Error

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}} = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{\nu}_i}{\frac{1}{n}\sum_{i}^{n}(X_i - \bar{X})^4}}$$

Remarks

OK, this is a bit nasty, but:

- It is less complicated than it seems. The numerator estimates var(v), the denominator estimates var(X).
- $SE(\hat{\beta}_1)$ is computed by regression software

Formula for $SE(\hat{\beta}_1)$

Standard Error

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}} = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{\nu}_i}{\frac{1}{n}\sum_{i}^{n}(X_i - \bar{X})^4}}$$

Remarks

OK, this is a bit nasty, but:

- It is less complicated than it seems. The numerator estimates var(v), the denominator estimates var(X).
- $SE(\hat{\beta}_1)$ is computed by regression software
- R has memorized this formula so you don't need to.

Summary: To test: $H0 : \beta_1 = \beta_{1,0}$ $vs.$ $H_1 : \beta_1 \neq \beta_{1,0}$

- Construct the t-statistics

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}}$$

- Reject at 5% significance level if $|t| > 1.96$
- The p-value is $p = Pr[|t| > |t^{act}|]$ = probability in tails of normal outside $|t^{act}|$;
  - you reject at the 5% significance level if the p-value is ¡ 0.05;
  - in general, you reject at the $\alpha \times 100\%$ significance level if the p-value is ¡ α;
- This procedure relies on the large-n approximation; typically $n = 50$ is large enough for the approximation to be excellent.

Confidence Intervals for β_1

Recall that a 95% confidence is, equivalently:

- The set of points that cannot be rejected at the 5% significance level;
- A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

Because the t-statistic for β_1 is $N(0, 1)$ in large samples, construction of a 95% confidence for β_1 is just like the case of the sample mean:

$$(\hat{\beta}_1 - 1.96 \times SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96 \times SE(\hat{\beta}_1))$$

Estimation, Testing and CI with R: testscore and str

- R command for linear regression is `lm`
- The basic usage is shown in the example below
- Notice that the standard output is rather basic

```
## Run the linear model function
lm(testscore ~ str, data = CASchools)
```

```
##
## Call:
## lm(formula = testscore ~ str, data = CASchools)
##
## Coefficients:
## (Intercept)          str
##      698.93        -2.28
```

Estimation, Testing and CI with R: testscore and str, ctd.

- We get more information if we use `summary`

```
summary(lm(testscore ~ str, data = CASchools))
```

```
##
## Call:
## lm(formula = testscore ~ str, data = CASchools)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -47.73 -14.25   0.48  12.82  48.54
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   698.93       9.47   73.82  < 2e-16 ***
## str            -2.28       0.48   -4.75  2.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19 on 418 degrees of freedom
## Multiple R-squared:  0.0512,Adjusted R-squared:  0.049
## F-statistic: 22.6 on 1 and 418 DF,  p-value: 2.78e-06
```

Confidence Interval

$$\textit{testscore} = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\ \textit{str}$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\right)$$

Confidence Interval

$$\textit{testscore} = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\,\textit{str}$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\right)$$

Confidence Interval

$$\underset{(9.47)}{testscore = 698.93} - \underset{(0.48)}{2.28}\ str$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\right) = \left(-2.28 \pm 1.96 \times 0.48\right) = (-3.22, -1.34)$$

- 90% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.64 \times SE(\hat{\beta}_1)\right)$$

Confidence Interval

$$\widehat{testscore} = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\right) = \left(-2.28 \pm 1.96 \times 0.48\right) = (-3.22, -1.34)$$

- 90% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.64 \times SE(\hat{\beta}_1)\right)$$

Confidence Interval

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)) = (-2.28 \pm 1.96 \times 0.48) = (-3.22, -1.34)$$

- 90% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.64 \times SE(\hat{\beta}_1)\right) = (-2.28 \pm 1.64 \times 0.48) = (-3.22, -1.49)$$

Hypothesis Testing

We test whether str has any effect on testscore

$$H0 : \beta_1 = 0 \; vs. \; H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

- The t-statistic is

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$

Hypothesis Testing

We test whether str has any effect on testscore

$$H0 : \beta_1 = 0 \ vs. \ H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

- The t-statistic is

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$

- Since $|t| > 1.64$, we **reject** the null hypothesis at 10%;

Hypothesis Testing

We test whether str has any effect on testscore

$$H0 : \beta_1 = 0 \ vs. \ H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

- The t-statistic is
$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$
- Since $|t| > 1.64$, we **reject** the null hypothesis at 10%;
- Since $|t| > 1.96$, we **reject** the null hypothesis at 5%;

Hypothesis Testing

We test whether str has any effect on testscore

$$H0 : \beta_1 = 0 \ vs. \ H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

- The t-statistic is
$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$
- Since $|t| > 1.64$, we **reject** the null hypothesis at 10%;
- Since $|t| > 1.96$, we **reject** the null hypothesis at 5%;
- Can we reject at 1.35%? Check and see whether the p-value is greater than 0.0135 ....

Hypothesis Testing

We test whether str has any effect on testscore

$$H0 : \beta_1 = 0 \ vs. \ H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

- The t-statistic is
$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$
- Since $|t| > 1.64$, we **reject** the null hypothesis at 10%;
- Since $|t| > 1.96$, we **reject** the null hypothesis at 5%;
- Can we reject at 1.35%? Check and see whether the p-value is greater than 0.0135 ....
- The p-value is $1.73e - 05$, thus we **reject** at any significance level.

p-value

- The p-value reported by R is the p-value for $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$;

Hypothesis Testing

p-value

- The p-value reported by R is the p-value for $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$;
- If your test is different, you can't use this p-value;

Hypothesis Testing

p-value

- The p-value reported by R is the p-value for $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$;
- If your test is different, you can't use this p-value;
- However, the test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ is the most common and has a special name: **significance test**

Hypothesis Testing

p-value

- The p-value reported by R is the p-value for $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$;
- If your test is different, you can't use this p-value;
- However, the test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ is the most common and has a special name: **significance test**
- If you **can't reject** the null hypothesis we say that the coefficient β_1 is **not** statistically significant.