Hypothesis testing and interval estimation in finite and large samples

Empirical Methods, Lecture 5

Tunga Kantarcı, FEB, Groningen University, Spring 2025

Hypothesis testing in finite and large samples

Consider the LRM

$$y_i = x_i\beta + \varepsilon_i.$$

So far we have been interested in estimating the β, the true effect in the population. Using the OLS method, we estimated it with $\hat{\beta}$.

We also learned that $\hat{\beta}$ is a random variable, has a sampling distribution, and hence a mean and a S.E.. This meant that some level of uncertainty is associated with the particular $\hat{\beta}$ due to that it is obtained by the random sample at hand.

Hypothesis testing is about testing whether β is equal to a certain hypothesized value, and we can use $\hat{\beta}$ to test that. But due to the random nature of $\hat{\beta}$, we have to take account of the uncertainty in $\hat{\beta}$ while conducting the test.

Hypothesis testing in finite and large samples

Let us hypothesise that the true β_k is equal to a particular β_k^0. Then, the null and the alternative hypotheses are

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0$$

Hypothesis testing in finite and large samples

The hypothesis we are testing is

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0.$$

That is, we want to check whether

$$\beta_k = \beta_k^0.$$

But we do not observe β_k. Hence, we cannot make this check. But we can estimate β_k. Suppose that $\hat{\beta}_k$ is the OLS estimate of β_k. Now we can check whether

$$\hat{\beta}_k = \beta_k^0.$$

Suppose that this is the case. Then, our test is complete, and we conclude that H_0 is true. But this conclusion has a problem.

Hypothesis testing in finite and large samples

$\hat{\beta}_k$ is a random variable, and it has a sampling distribution. Hence, there is a probability associated with the condition

$$\hat{\beta}_k = \beta_k^0.$$

Therefore, we need to check if the equality holds in a statistical sense. We do this check by constructing a test statistic based on the random variable $\hat{\beta}_k$. But since $\hat{\beta}_k$ is random, the test statistic becomes a random variable.

Hypothesis testing in finite and large samples

A test statistic is a random variable, and it has a distribution. What determines the distribution of, e.g., the t statistic?

As we will se later in these slides, the t statistic is a function of $\hat{\boldsymbol{\beta}}$, and if we take $\boldsymbol{X}$ as given, it is a function of $\boldsymbol{\varepsilon}$. Hence, the distribution of $\boldsymbol{\varepsilon}$ determines the distribution of the t statistic.

Hypothesis testing in finite and large samples

If we assume that ε is normal, the test statistic has an exact distribution. An exact distribution means that the distribution is valid for any finite n. For example, the t statistic has a t distribution if ε is normal. This is a distribution tabulated at the back of textbooks.

Hypothesis testing in finite and large samples

If ε is not normal, the test statistic does not have an exact distribution. But if we require that n is large, then the test statistic has an asymptotic distribution that approximates an exact distribution, in particular the normal distribution.

Hypothesis testing in finite samples: single restriction

Recall that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}.$$

We know from the previous lecture that if

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

then

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

This means that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Hypothesis testing in finite samples: single restriction

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Often we are interested in testing a linear restriction on a given true coefficient. E.g.,

$$\begin{aligned} H_0 &: \beta_k = \beta_k^0 \\ H_1 &: \beta_k \neq \beta_k^0. \end{aligned}$$

That is, we are interested in the element located in the row k of the $K \times 1$ vector $\boldsymbol{\beta}$, and in the observation located in the row k and column k of the $K \times K$ matrix $\boldsymbol{X}'\boldsymbol{X}$. That is, we are interested in

$$\hat{\beta}_k - \beta_k \mid \boldsymbol{X} \sim N\left[0, \sigma^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}\right].$$

Define

$$S^{kk} \equiv \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}.$$

Hypothesis testing in finite samples: single restriction

$$\hat{\beta}_k - \beta_k \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right].$$

Under the null hypothesis, we have $\beta_k = \beta_k^0$. Hence,

$$\hat{\beta}_k - \beta_k^0 \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right],$$

assuming that $\mathrm{E}\left[\hat{\beta}_k \mid \boldsymbol{X}\right] = \beta_k^0$.

We want $\hat{\beta}_k - \beta_k^0$ because we want to test whether this is equal to 0.

Hypothesis testing in finite samples: single restriction

$$\hat{\beta}_k - \beta_k^0 \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right].$$

Standardise $\hat{\beta}_k - \beta_k^0$ to get

$$z_k \mid \boldsymbol{X} \equiv \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \,\Bigg|\, \boldsymbol{X} \sim N\left[0, 1\right].$$

The distribution does not depend on $\hat{\beta}_k$, β_k^0, σ, or $\boldsymbol{X}$. Hence,

$$z_k \sim N\left[0, 1\right].$$

This is a convenient simplification because we do not need to condition on $\boldsymbol{X}$ while using the test statistic.

Hypothesis testing in finite samples: single restriction

$$z_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}.$$

z_k is not usable because σ^2 is unknown. Replace σ^2 with its unbiased estimator

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}.$$

We obtain

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

But we change from z_k to t_k. So how is t_k distributed?

Hypothesis testing in finite samples: single restriction

The only difference between

$$z_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \sim N[0,1]$$

and

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}} \sim t[n-K]$$

is that we have replaced the unknown σ^2 by its estimator $\hat{\sigma}^2$, and the result is that we move from a standard normal distribution to a t distribution which has slightly thicker tails.

Hypothesis testing in finite samples: single restriction

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

What is the intuition of this test statistic?

Is the distance between $\hat{\beta}_k$ and β_k^0 sufficiently large, with the distance measured in terms of the sampling variance of $\hat{\beta}_k$? Is t_k sufficiently large? If it is, reject the null, for the true β_k, that $\beta_k = \beta_k^0$. This is the decision rule of the test.

Hypothesis testing in finite samples: single restriction

How large t_k should be depends on the threshold t value we want to consider. We call this threshold t value, the critical t value, and denote it as

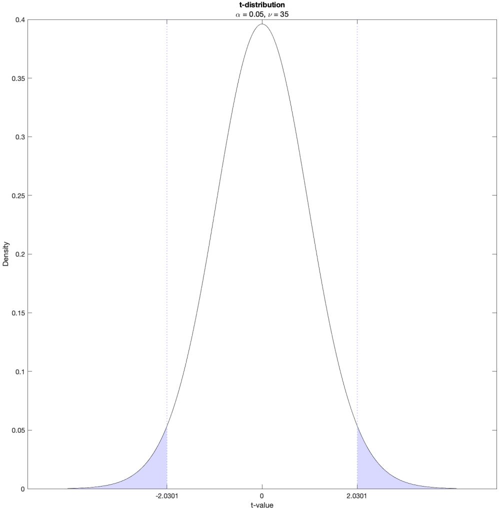
$$t^c_{\frac{\alpha}{2}, n-K}.$$

t^c is a value from the t distribution and depends on

- $\frac{\alpha}{2}$: area under the t distribution covering up to where we want t^c to rest. Hence, $\frac{\alpha}{2}$ determines the t^c we want to consider. It is often taken as 5%. So "statisticians are people whose aim in life is to be wrong 5% of the time" (Kempthorne and Doerfler, 1969).
- $n - K$: degrees of freedom which determines the shape of the t distribution.

Hypothesis testing in finite samples: single restriction

For example, for $\alpha = 0.05$ and $n - K = 35$, we have



Hypothesis testing in finite samples: single restriction

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}.$$

One point left unclear is that the distance $\hat{\beta}_k - \beta_k^0$ can be positive or negative. Hence, t_k can be positive or negative.

If t_k is positive, we reject the null if

$$t_k > t^c_{1-\frac{\alpha}{2},n-K}.$$

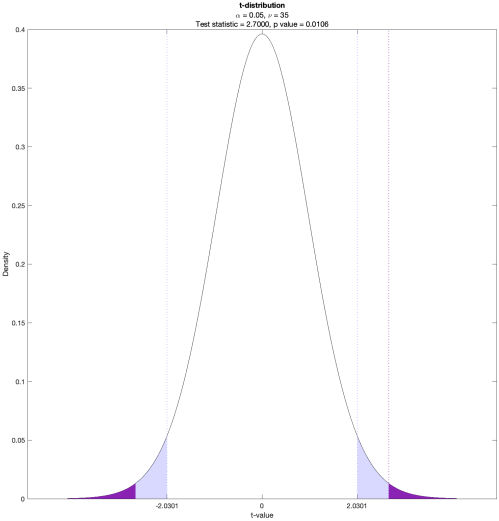
If t_k is negative, we reject the null if

$$t_k < t^c_{\frac{\alpha}{2},n-K}.$$

This is what is called a two-tailed t test.

Hypothesis testing in finite samples: single restriction

For example, for $\alpha = 0.05$, $n - K = 35$, and $t_k = 2.7$, we have



Hypothesis testing in finite samples: single res. example

Mincer (1974) considers the regression of the log of wage on exper, educ, and IQ score. The data contains 935 observations. Estimation of this regression gives

$$\hat{\beta}_{IQ} = 0.0058$$

with

$$\text{Est. S.E.}\left[\hat{\beta}_{IQ} \mid \boldsymbol{X}\right] = \sqrt{\hat{\sigma}^2 S^{IQ}} = 0.001.$$

Hypothesis testing in finite samples: single res. example

Someone claims that each additional IQ point raises one's wage by 0.0075 on average. That is,

$$\beta_{IQ}^{0} = 0.0075.$$

We want to test this claim. The null and the alternative are

$$\begin{aligned} H_0 &: \beta_{IQ} = 0.0075 \\ H_1 &: \beta_{IQ} \neq 0.0075. \end{aligned}$$

A two-tailed test!

Hypothesis testing in finite samples: single res. example

We need to calculate the t and t^c, and compare t to t^c to decide on the result of the test.

Hypothesis testing in finite samples: single res. example

t is calculated as follows.

$$t = \frac{0.0058 - 0.0075}{0.001} = -1.75.$$

Hypothesis testing in finite samples: single res. example

t^c is calculated as follows.

Consider a significance level of 0.05. Then, for this two-tailed test, $\frac{\alpha}{2} = 0.025$.

The degrees of freedom is $935 - 4 = 931$.

Then,

$$t^c_{0.025,931} = -1.9625,$$

using the tabulated t distribution at the back of your textbook.

Hypothesis testing in finite samples: single res. example

Since

$$t > t^c,$$

that is, since

$$-1.7500 > -1.9625,$$

we fail to reject the null hypothesis.

Hypothesis testing in finite samples: single res. example

We can also compare p to p^c to decide on result of the t test.

p is the p value corresponding to the t value.

p^c is the critical p value corresponding to t^c, the critical t value.
p^c is what we call the significance level.

Hypothesis testing in finite samples: single res. example

p is calculated, for this two-tailed test, as

$$p = 2 * p_{-1.75,931} = 0.0805$$

using standard statistical software. The tabulated t distribution at the back of your textbook will not present this exact number because tabulations cannot be too detailed since there is no space to present them.

Hypothesis testing in finite samples: single res. example

p^c is calculated as

$$p^c = 2 * p_{-1.9625,931} = 0.05$$

using standard statistical software, or the tabulated t distribution at the back of your textbook will present this number because 0.05 is a conventional critical level.

Hypothesis testing in finite samples: single res. example

Since

$$p > p^c,$$

that is, since

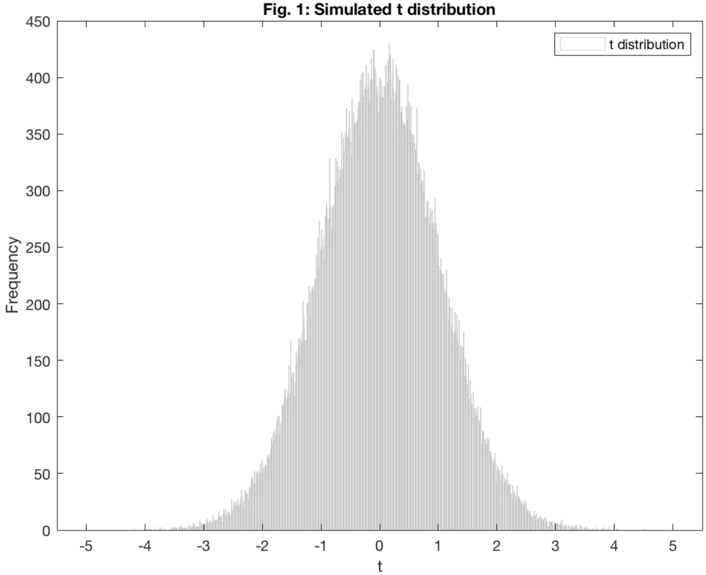
$$0.0805 > 0.0500,$$

we fail to reject the null hypothesis.

Hypothesis testing in finite samples: single res. example

O'Hara (2018) proposes that econometrics instructors move away from using the tabulated distribution of the test statistic at the back of the textbooks when teaching hypothesis testing. Instead, he proposes that instructors teach students to test hypotheses by using the simulated distribution of the test statistic which can be created using random number generators in statistical software. This provides students with a visual and intuitive understanding of the sampling distribution and the logic behind hypothesis testing. In the next few slides we will follow what O'Hara proposes.

Hypothesis testing in finite samples: single res. example

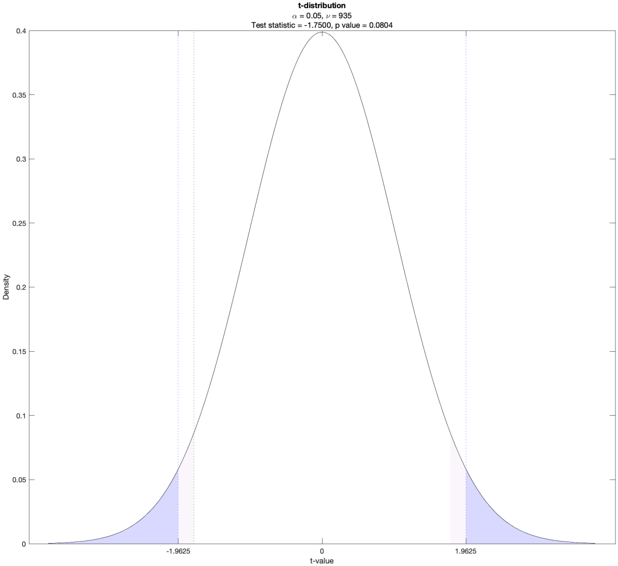


Hypothesis testing in finite samples: single res. example

In this plot, a probability area – p or p^c – represents the fraction that the t values occur up to some t value – t or t^c – in all t values in the distribution.

Now that we understand what we are doing, we can replace the simulated frequency distribution with the continuous PDF of the t statistic.

Hypothesis testing in finite samples: single res. example



Hypothesis testing in large samples

If

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

holds, the exact sampling distribution of $\hat{\boldsymbol{\beta}}$, conditional on $\boldsymbol{X}$, is

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

When $\hat{\boldsymbol{\beta}}$ is normal, the t statistic has the exact t distribution. This is what we have shown above.

Hypothesis testing in large samples

If

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

does not hold, the t statistic does not have the exact t distribution in finite n. The same holds for the F statistic.

What happens then?

Consider the t statistic.

Hypothesis testing in large samples: single restriction

$$
\begin{aligned}
t_k &= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 \left[\left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \right]_{k,k}}} \frac{\sqrt{n}}{\sqrt{n}} \\
&= \frac{\sqrt{n} \left(\hat{\beta}_k - \beta_k^0 \right)}{\sqrt{\hat{\sigma}^2 \left[\left(\frac{1}{n} \boldsymbol{X}'\boldsymbol{X} \right)^{-1} \right]_{k,k}}}.
\end{aligned}
$$

Hypothesis testing in large samples: single restriction

Consider the numerator of

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}}.$$

The derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}$ shows that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right].$$

In this derivation we did not assume that ε is normal! The normal distribution is due to the CLT! Considering the element k of $\hat{\boldsymbol{\beta}}$, and that under the null $\beta_k = \beta_k^0$,

$$\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right) \xrightarrow{d} N\left[0, \sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}\right].$$

Hypothesis testing in large samples: single restriction

Consider the denominator of

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}}.$$

It can be shown that

$$\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}} \xrightarrow{d} \sqrt{\sigma^2\left[\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i'\right]\right)^{-1}\right]_{k,k}}.$$

Using the ratio rule of limiting distributions,

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2\left[\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}} \xrightarrow{d} N\left[0,1\right].$$

Hypothesis testing in large samples: single restriction

Drop the two instances of $\sqrt{n}$. We have

$$t_k = \frac{\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{\hat{\sigma}^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}}} \xrightarrow{d} N\left[0,1\right].$$

So

$$t_k \overset{a}{\sim} N\left[0,1\right].$$

This shows that the t statistic approximately has a standard normal distribution in finite but large samples. Hence, if n is large, we can compare the t statistic with the critical values from a standard normal distribution. We do not need to assume that ε is normal!

Power of a test statistic

A test is said to have good power if the probability of rejecting the null hypothesis, when it is false, is high.

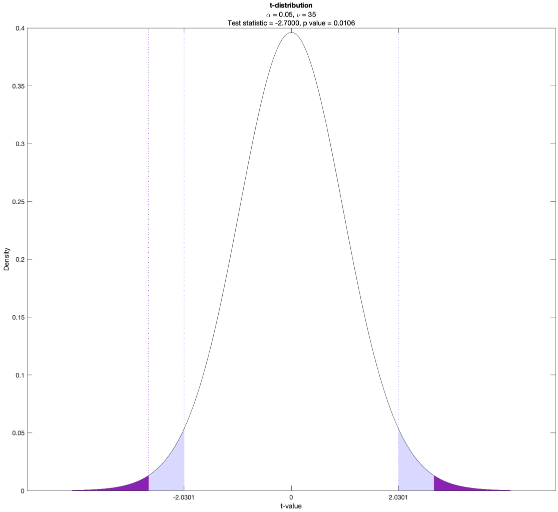
Consider the null and alternative hypotheses

$$H_0 : \beta_k = \beta_k^0$$
$$H_1 : \beta_k \neq \beta_k^0.$$

When is the null hypothesis false? If the alternative is true. When is the alternative true? If you are far on the left or right hand side of the t distribution.

Power of a test statistic

E.g., suppose $\beta_k^0 = 0$. If $-|t_k|$ is smaller than $-|t^c|$, or if $|t_k|$ larger than $|t^c|$, we reject the null.



Power of a test statistic

When are you far on the left or right hand side of the, say, t distribution? In three situations.

Power of a test statistic

First, if the alternative is true. This happens when the effect size is large. Effect size refers to the size of β_k. If β_k is large, then the alternative hypothesis is more likely. β_k is unobserved, but if β_k is large, then the sample we observe is likely to reflect this, and we will have a large $\hat{\beta}_k$, and consequently a large t value.

Power of a test statistic

Second, it happens when the sample size is large. If the sample size is large, the S.E. of the coefficient estimate is smaller, and consequently the t has a larger value.

Power of a test statistic

Third, if α is larger, because this ensures that you are automatically further on the left or on the right of the t distribution. Since everybody agrees on an α value of 5%, this factor is hardly relevant.

Interval estimation

Is the true coefficient β_k equal to a certain value (β_k^0)?

To answer this question, utilizing the sample data at hand, we have point estimated β_k using the OLS method, and developed a test statistic to check how close $\hat{\beta}_k$ and β_k^0 are in a statistical sense.

Interval estimation

Can we estimate a lower and upper bound for the true coefficient β_k?

To answer this question, utilizing the sample data at hand, we can construct an interval estimate.

Interval estimation

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

Interval estimation

We know that, in a finite sample,

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}} \sim t\left[n - K\right]$$

where

$$s_{\hat{\beta}_k} \equiv \sqrt{\hat{\sigma}^2 S^{kk}}.$$

Interval estimation

Then, we can state that

$$\text{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}} < t_{\alpha/2,\nu}\right) = 1 - \alpha,$$

where

- α is some probability value, and
- $-t_{\alpha/2}$ and $t_{\alpha/2}$ are some lower and upper thresholds, or critical values as we have seen.

Interval estimation

Interpret

$$\text{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}} < t_{\alpha/2,\nu}\right) = 1 - \alpha.$$

The probability that the random variable

$$\frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}}$$

is between the stated thresholds is $1 - \alpha$.

Interval estimation

For example, if $\nu = 999$ and $\alpha = 0.05$,

$$t_{0.025,999} = 1.9623,$$

and hence

$$\text{Prob}\left(-1.9623 < \frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}} < 1.9623\right) = 0.95.$$

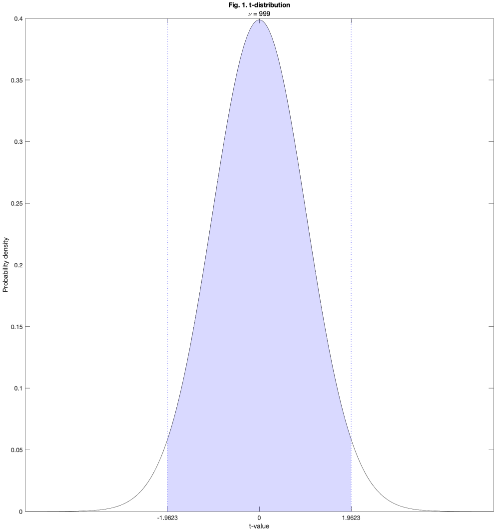
The probability that the random variable

$$\frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}}$$

is between the stated boundaries is 0.95.

Interval estimation

The shaded area between the stated thresholds is 0.95.



Interval estimation

Now rearrange the terms of

$$\text{Prob}\left(-t_{\alpha/2,\nu} < \frac{\hat{\beta}_k - \beta_k^0}{s_{\hat{\beta}_k}} < t_{\alpha/2,\nu}\right) = 1 - \alpha,$$

to obtain

$$\text{Prob}\left(\hat{\beta}_k - t_{\alpha/2} s_{\hat{\beta}_k} < \beta_k^0 < \hat{\beta}_k + t_{\alpha/2} s_{\hat{\beta}_k}\right) = 1 - \alpha.$$

Interval estimation

At this instance the interpretation changes.

$$\text{Prob}\left(\hat{\beta}_k - t_{\alpha/2} s_{\hat{\beta}_k} < \beta_k^0 < \hat{\beta}_k + t_{\alpha/2} s_{\hat{\beta}_k}\right) = 1 - \alpha.$$

Notice two things.

Interval estimation

$$\text{Prob}\left(\hat{\beta}_k - t_{\alpha/2}s_{\hat{\beta}_k} < \beta_k^0 < \hat{\beta}_k + t_{\alpha/2}s_{\hat{\beta}_k}\right) = 1 - \alpha.$$

First, the interpretation is for the unique nonrandom population parameter β_k^0.

Interval estimation

$$\text{Prob}\left(\hat{\beta}_k - t_{\alpha/2} s_{\hat{\beta}_k} < \beta_k^0 < \hat{\beta}_k + t_{\alpha/2} s_{\hat{\beta}_k}\right) = 1 - \alpha.$$

Second, the end points of the interval are random because $\hat{\beta}_k$ is random. $\hat{\beta}_k$ has a sampling distribution. We are taking samples from the population repeatedly, and estimating an interval using each sample. Hence, we have a series of estimated intervals resulting from repeated sampling. But since we are not able to do repeated sampling, we are bound to use one estimate of an interval using the data at hand.

Interval estimation

$$\text{Prob}\left(\hat{\beta}_k - t_{\alpha/2} s_{\hat{\beta}_k} < \beta_k^0 < \hat{\beta}_k + t_{\alpha/2} s_{\hat{\beta}_k}\right) = 1 - \alpha.$$

Then, the interpretation is as follows. In repeated sampling, the true population parameter β_k^0 falls within intervals, like the one we estimated using the data at hand, $1 - \alpha$ of the times.

Interval estimation

Given the single sample at hand, we have only one estimate of the interval. The probability that the interval we estimate using the data at hand contains β_k^0 is either 0 or 1. Hence, it is incorrect to say that the probability that the interval we estimated using the data at hand contains β_k^0 is 95 percent. The interval we calculated is just an estimate of one of the many intervals that contain β_k^0 95 percent of the times.

Interval estimation

Building on the earlier example, if $\nu = 999$ and $\alpha = 0.05$,

$$t_{0.025,999} = 1.9623.$$

Suppose

$$\hat{\beta}_k = 0.4574$$

and

$$s_{\hat{\beta}_k} = 0.0557.$$

We get

$$\text{Prob}\left(0.3482 < \beta_k^0 < 0.5667\right) = 95\%.$$

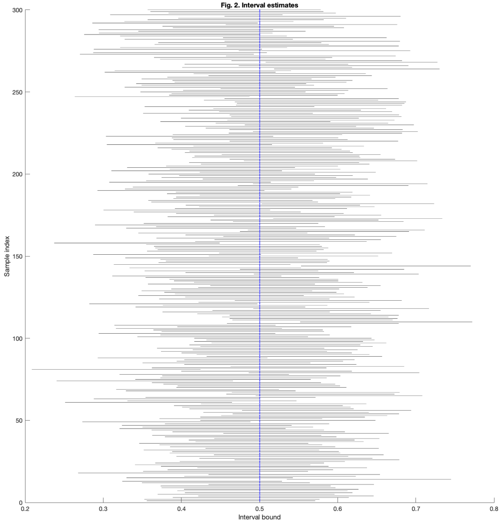
So an interval estimate using the sample data at hand is

$$[0.3482, 0.5667].$$

This is called a "confidence" interval because we use this, and only one interval, to be confident about the population coefficient to a certain probability extent.

Interval estimation

β_k^0 falls within all intervals 95 percent of the times:



Interval estimation, example

A test and a confidence interval are closely related.

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

We reject the null $\beta_{educ} = 0$ of the t test since it lies outside the confidence interval.

Hypothesis testing in finite samples: multiple restrictions

We might want to hypothesize that there are J linear restrictions on the true coefficient vector $\boldsymbol{\beta}$ against alternatives such that

$$H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{q}$$
$$H_1 : \boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{q}.$$

$\boldsymbol{R}$ is a matrix of J restrictions for K parameters. $J \times K$. $\boldsymbol{\beta}$ is the true coefficient vector. $K \times 1$. $\boldsymbol{q}$ is the hypothesized value of $\boldsymbol{R}\boldsymbol{\beta}$. $J \times 1$. E.g.,

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{R}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\boldsymbol{q}},$$

implies the linear restrictions

$$\beta_1 + \beta_2 = 0,$$
$$\beta_2 = 1.$$

Hypothesis testing in finite samples: multiple restrictions

We want to test whether our hypothesis is true. We do not observe $\boldsymbol{\beta}$, but we can estimate it with $\hat{\boldsymbol{\beta}}$. Suppose

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}.$$

We could conclude that H_0 is true. But this conclusion has a problem.

Hypothesis testing in finite samples: multiple restrictions

Remember that $\hat{\boldsymbol{\beta}}$ is a random variable and has a sampling distribution. Hence, there is a probability associated with the condition

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}.$$

Therefore, we need to check if

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$$

holds in statistical terms. We do this check through a test statistic based on the random variable

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}.$$

Hypothesis testing in finite samples: multiple restrictions

$\boldsymbol{R}\hat{\beta} - \boldsymbol{q}$ is a random variable, and therefore it has a distribution.
Therefore, we start by studying this distribution.

Hypothesis testing in finite samples: multiple restrictions

Taking the expectation conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q} \mid \boldsymbol{X}\right] &= \mathrm{E}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] - \mathrm{E}\left[\boldsymbol{q} \mid \boldsymbol{X}\right] \\
&= \boldsymbol{R}\mathrm{E}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] - \boldsymbol{q} \\
&= \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{q} \\
&= \mathbf{0},
\end{aligned}$$

assuming that

$$\mathrm{E}\left[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right] = \boldsymbol{\beta},$$

and under the null

$$\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{q}.$$

Hypothesis testing in finite samples: multiple restrictions

Taking the variance conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\text{Var}\left[\boldsymbol{R}\hat{\beta} - \boldsymbol{q} \mid \boldsymbol{X}\right] &= \text{Var}\left[\boldsymbol{R}\hat{\beta} \mid \boldsymbol{X}\right] \\
&= \boldsymbol{R}\text{Var}\left[\hat{\beta} \mid \boldsymbol{X}\right]\boldsymbol{R}' \\
&= \boldsymbol{R}\sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}' \\
&= \sigma^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'.
\end{aligned}$$

Hypothesis testing in finite samples: multiple restrictions

We know that if

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{I}\right]$$

then

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Since $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}$ is a linear function of $\hat{\boldsymbol{\beta}}$,

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q} \mid \boldsymbol{X} \sim N\left[\mathbf{0}, \sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right],$$

with the mean and variance derived above.

Hypothesis testing in finite samples: multiple restrictions

It can be shown that

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\boldsymbol{R}\hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)}{J} \sim F\left[J, n-K\right],$$

where J and $n - K$ are the numerator and denominator degrees of freedom. These two parameters result from the derivation of the F statistic which is not shown here.

Hypothesis testing in finite samples: multiple restrictions

$$F = \frac{\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)' \left[\boldsymbol{R}\hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}\right)}{J}.$$

What is the intuition of the test statistic? Is the distance between $\boldsymbol{R}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{q}$ sufficiently large, with the distance measured in terms of the sampling variance of $\boldsymbol{R}\hat{\boldsymbol{\beta}}$? Is F sufficiently large? If it is, reject the null that $\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$. This is the decision rule of the test.

Hypothesis testing in finite samples: multiple restrictions

How large should F be depends on the threshold F value we want to consider. This threshold, or critical, F value is

$$F^c_{1-\alpha,J,n-K},$$

where

- F^c : a value from the F distribution which depends on the following,
- $1-\alpha$: area under the F distribution covering up to where we want F^c to rest. Hence, $1-\alpha$ determines the threshold F^c we want to consider,
- J, $n-K$: two degrees of freedom parameters as arguments of the F distribution.

Hypothesis testing in finite samples: multiple restrictions

This means that we would reject the null hypothesis if

$$F > F^{c}_{1-\alpha,J,n-K}.$$

If the null is rejected, we conclude that the restrictions we impose on the parameters in the null hypothesis are jointly not significant.

The test does not inform about which restriction is not significant: any or all restrictions are not significant.

Hypothesis testing in finite samples: multiple res. example

Mincer (1974) estimates a regression of the log of wage on a constant term, work experience, education (in years), and IQ score. The data contains 935 observations. We add to this regression a quadratic function of age, that is, we add age and age^2.

Hypothesis testing in finite samples: multiple res. example

Someone claims that age has no effect on wage. That is,

$$\beta^0_{age} = 0,\ \beta^0_{age^2} = 0.$$

We want to test this claim. The null and the alternative hypotheses are

$$H_0 : \boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{q}$$
$$H_1 : \boldsymbol{R}\hat{\boldsymbol{\beta}} \neq \boldsymbol{q}.$$

Checkpoint. Note that this is a one-tailed test.

Hypothesis testing in finite samples: multiple res. example

We need to calculate F, and compare it to F^c to decide on the result of the test.

Hypothesis testing in finite samples: multiple res. example

With additional information on $\hat{\sigma}$, the F statistic can be calculated as described above. It turns out that

$$F = 4.5735.$$

Hypothesis testing in finite samples: multiple res. example

Consider a significance level of 0.05. Hence, for this one-tailed test, $\alpha = 0.05$. The numerator degrees of freedom, J, is 2, and the denominator degrees of freedom, $n - K$, is $935 - 6 = 929$. F^c can then be calculated as

$$F^c = 3.0054,$$

using the tabulated F distribution at the back of your textbook, or using statistical software.

Hypothesis testing in finite samples: multiple res. example

Since

$$F > F^c,$$

that is, since

$$4.5735 > 3.0054,$$

we reject the null hypothesis.

Hypothesis testing in finite samples: multiple res. example

We can also compare p to p^c to decide on result of the F test. p is the p value corresponding to the F value. p^c is the critical p value corresponding to the critical F value.

Hypothesis testing in finite samples: multiple res. example

p can be calculated as

$$p = p_{4.5735,2,929} = 0.0106$$

using standard statistical software. The tabulated F distribution at the back of your textbook will not present this number because tabulations cannot be too detailed for space reasons.

Hypothesis testing in finite samples: multiple res. example

p^c can be calculated as

$$p^c = p_{3.0054,2,929} = 0.0500$$

using standard statistical software, and the tabulated F distribution at the back of your textbook will present this number because 0.05 is a conventional critical level.

Hypothesis testing in finite samples: multiple res. example

Since

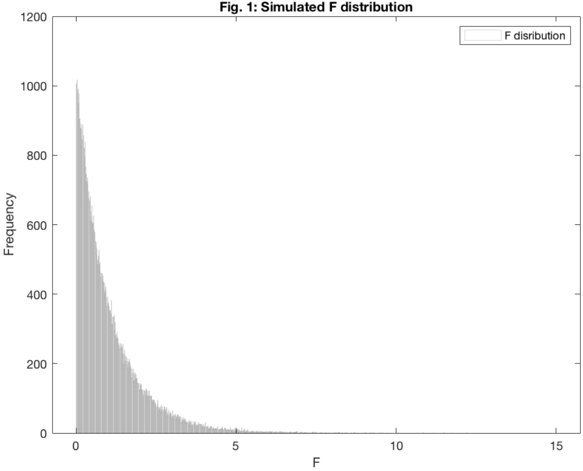
$$p > p^c,$$

that is, since

$$0.0106 < 0.0500,$$

we reject the null hypothesis.

Hypothesis testing in finite samples: multiple res. example



Hypothesis testing in finite samples: multiple res. example