Statistical inference

Edicson Luna

University of Pennsylvania

Sampling distribution of the OLS estimators

T-test of the estimators

F-test

Sampling distribution of the OLS estimators

Normality

We now add a new assumption (in addition to the 5 we studied before).

- **Assumption 6.** The error ϵ_i is independent of X (all the set of covariates) and is normally distributed with zero mean and variance σ^2.

$$\epsilon_i | x_i \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma^2)$$

Notice that this assumption implies assumptions 4 and 5 (hence, they become redundant). This assumption (6) will help us with the following theorem, but it is unnecessary if n is large. We can also say $\epsilon | X \sim \mathcal{N}(0, \sigma^2 I_n)$.

CLM

This set of 6 assumptions is called the **classical linear model (CLM) assumptions**. A way to summarize the list of 6 assumptions is to say

$$y_i|x_i \sim_{\text{i.i.d.}} \mathcal{N}(\beta_0 + \beta_1 x_{i1} + ... + \epsilon_i, \sigma^2)$$

where there is no multicollinearity among the X (the complete matrix of independent variables).

Normal sampling distribution

Theorem. Under the CLM assumptions (1 to 6) conditional on X

$$\hat{\beta}|X \sim \mathcal{N}(\beta, \text{Var}(\hat{\beta}|X))$$

Where $\text{Var}(\hat{\beta}|X) = (X'X)^{-1}\sigma^2$. Notice that, for each coefficient j, this is the same as saying,

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\text{Var}(\hat{\beta}_j \mid X)}} \mid X \sim \mathcal{N}(0, 1),$$

where $\text{Var}(\hat{\beta}_j \mid X) = \sigma^2 \left[(X'X)^{-1}\right]_{jj}$.

- Under the CLM assumptions (even without A6), $\hat{\beta}$ is the best linear unbiased estimator (BLUE). It is the "best" in the sense that minimizes the variance.

T-test of the estimators

t distribution

Theorem. Under the CLM assumptions (1 to 6), for each coefficient β_j in β, $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j|X)} \mid X \sim t_{n-k-1}$, where $se(\hat{\beta}_j|X) = \sqrt{\hat{\sigma}^2 \left[(X'X)^{-1}\right]_{jj}}$, and $\hat{\sigma}^2 = \frac{SSR}{n-k-1}$.

The term t_{n-k-1} denotes a t distribution with $n-k-1$ degrees of freedom. Here, $k+1$ is the number of estimated parameters in the model, including the intercept. Therefore, the degrees of freedom are $n-(k+1) = n-k-1$.

This result is similar to the previous normality result, but now we replace the unknown σ^2 with its estimator $\hat{\sigma}^2$. This extra estimation uncertainty is why the statistic follows a t distribution instead of a standard normal distribution.

Testing

Usually, we are interested in testing a hypothesis of the form

$$\mathrm{H}_0 : \beta_j = 0$$

This will imply $\frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j|X)} \sim \mathrm{t}_{n-k-1}$.

We will now analyze how to conclude if we should reject or not reject H_0.

One-sided alternatives

Consider a case in which it is only reasonable to have a one-sided alternative hypothesis

$$H_1 : \beta_j > 0$$

Here, we could redefine $\mathrm{H}_0 : \beta_j \leq 0$, and analyze the distribution only on the right side.

Notice that the rule we had before of $\Phi(c) = 1 - \frac{\alpha}{2}$ does not apply because we are analyzing only one tail. In this case, the c for rejection should be found by $\Phi(c) = 1 - \alpha$ ($\Phi(c) = \alpha$ if the alternative hypothesis is left-sided).

Two-sided alternatives

In the we have two-sided alternatives we go back to $\mathrm{H}_0 : \beta_j = 0$ and then

$$\mathrm{H}_1 : \beta_j \neq 0$$

In this case, we go back to the type of hypothesis tests we studied in the first sessions where we worked with $c = 1.96$ to reject the null hypothesis.

Testing other hypothesis

In general, we are interested in studying if a coefficient is different than zero, but sometimes we may want to analyze if it's different than a_j (a real value). The analysis is the same, it just will define

$$\frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim \mathrm{t}_{n-k-1}$$

and then proceed with the same procedure for hypothesis testing.

P-Values

Remember that the P-value is the threshold of the t-statistic to just reject the null hypothesis. In other words, it is that α value where we both accept and reject.

Imagine that you get a t-statistic value of 1.85. Here, you can not reject the null hypothesis for $\alpha = 0.05$. Nonetheless, if we define $\alpha = 0.1$, we would reject the null hypothesis ($c = 1.645 < 1.85$). In this case, the minimum α in which we reject the null hypothesis is 0.072. This is its respective P-Value.

Interpretation

We prefer to say "we fail to reject H_0 at the α level", instead of saying "we accept H_0 at the α level".

The reason for this is that there are many values that we may not reject. If we just say "we accept...", we would be accepting a large amount of null hypothesis, which does not make sense.

An important consideration is that β is our parameter of interest (not the significance level). If we find that the result is statistically significant, but the β is too low, the result is irrelevant. And vice versa.

Confidence intervals

This part is the same as before. We define a confidence interval for the parameter $[\hat{A}_n, \hat{B}_n]$ where

$$A_n = \hat{\beta}_j - c \cdot \text{se}(\hat{\beta}_j)$$

$$B_n = \hat{\beta}_j + c \cdot \text{se}(\hat{\beta}_j)$$

Where c depends on the α we define (1.96 for $\alpha = 0.05$).

Linear combination of parameters

What if we want to analyze a hypothesis of the form

$$\mathrm{H}_0 : \beta_1 = \beta_2$$

The procedure is much simpler than you may think. Just define $\theta = \beta_1 - \beta_2$ and the new hypothesis is $\mathrm{H}_0 : \theta = 0$. As long as we can find $\mathrm{se}(\theta)$ we can apply what we did before.

F-test

Motivation

Sometimes we are interested in testing a null hypothesis including several variables. For instance, consider

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \quad \textbf{(Model 1)}$$

And we want to estimate the null hypothesis

$$\mathrm{H}_0 : \beta_1 = 0 \wedge \beta_2 = 0$$

$$\mathrm{H}_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0$$

How should we proceed?

SSR

Notice that we are tempted to consider two different t-tests and reject if one is rejected.

There is a better way to carry out this test. But first, remember our definition of the residual sum of squares

$$\mathrm{SS}_{\mathrm{residual}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e^2$$

Remember that this sum of residuals is inversely related to the number of variables. The more variables you add to the model, the smaller the $\mathrm{SS}_{\mathrm{residual}}$.

Intuition

Then, we are trying to study whether the inclusion of X_{i1} and X_{i2} in the model are successfully explaining the data. That is, if the SS_{residual} is being reduced sufficiently enough. To do this, we need to analyze the model without those two variables

$$Y_i = \beta_0 + \beta_3 X_{i3} + \epsilon_i \quad \textbf{(Model 2)}$$

And we compute the SS_{residual} to compare it with the initial model.

Formulation

Call the $\mathrm{SS}_{\mathrm{residual}}$ from model 1 "SSR^1" and the $\mathrm{SS}_{\mathrm{residual}}$ from model 2 "SSR^2". Then define the **F statistic** as

$$F := \frac{(\mathrm{SSR}^2 - \mathrm{SSR}^1)/2}{\mathrm{SSR}^1/(n-k-1)}$$

The 2 comes from the number of parameters we restrict in the model. This statistic under the null hypothesis is distributed as $F \sim \mathbf{F}_{2,n-k-1}$.

Generalization

In general, call SSR^{ur} to the unrestricted model (the one with all the variables) and SSR^{r} to the restricted model. Also, say q is the number of restrictions (the number of coefficients equalized to 0 in our null hypothesis). Then,

$$F := \frac{(\text{SSR}^{\text{r}} - \text{SSR}^{\text{ur}})/2}{\text{SSR}^{\text{ur}}/(n-k-1)}$$

Where

$$F \sim \mathbf{F}_{q,n-k-1}$$

Rejection of the null hypothesis

With our F statistic, we may analyze if it is greater than a critical value c for a given α. The critical value c depends on q, $n-k-1$ and α. Then, there are no general values such as the 1.96 in the t-statistic for a large sample.

Other expressions for F

F can also be defined using the R^2 of the restricted and unrestricted models.

$$F := \frac{(R^2_{\text{ur}} - R^2_{\text{r}})/q}{(1 - R^2_{\text{ur}})/(n-k-1)}$$

Notice that an overall significance would give an R^2 of zero. Then, if we want to test all the variables at the same time, we work with

$$F := \frac{R^2/k}{(1 - R^2)/(n-k-1)}$$