

Statistical Inference

Overview

This note reviews material that should be familiar from undergraduate texts. However, we will borrow the notation from Handouts 1-4 to make it relevant to this module.

The note reviews the following topics:

- confidence intervals;
- simple hypothesis tests;
- p-values;
- statistical power.

It does not cover (multiple) linear hypotheses, which is the subject of Handout 4.

Set-up

Consider the model,

$$Y = \beta_1 X_1 + X_2 \beta_2 + u$$

where β_1 is a scalar. We know that under CLRM 1-6,¹

$$\hat{\beta}_1 | X \sim N(\beta_1, \sigma^2 / (X_1' M_2 X_1))$$

We can write $\sigma^2 (X_1' M_2 X_1)^{-1}$ as a fraction in this way, since $\hat{\beta}_1$ is a scalar.

Let us define,

$$V = \sigma^2 / (X_1' M_2 X_1) \\ \text{and } \hat{V} = s^2 / (X_1' M_2 X_1)$$

Where the difference is that in \hat{V} includes, s^2 , the estimator for σ^2 . Thus,

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{V}} \sim N(0, 1) \\ \text{and } \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}}} \sim T_{n-k}$$

¹In Handout 4 we showed that β_1 can be written as $c' \beta$ where $c = [1 \ 0 \ \dots \ 0]'$. Therefore, it must be that $(X_1' M_2 X_1)^{-1} = c' (X' X)^{-1} c$

Confidence interval

Here we are going to describe the confidence interval (CI) for a single estimator. You can expand this to the vector case, but must then consider the joint distribution of the estimators.

Given the unknown population parameter, β_1 , we want to define the CI such that,

$$Pr(\beta_1 \in CI_{1-\alpha}|X) = 1 - \alpha$$

We know that $(\hat{\beta}_1 - \beta_1)/\sqrt{V} \sim N(0, 1)$. Therefore, using the percentiles of the standard normal distribution, we can say that,

$$\begin{aligned} 1 - \alpha &= Pr\left(z_{\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V}} \leq z_{1-\alpha/2}\right) \\ &= Pr(z_{\alpha/2}\sqrt{V} \leq \hat{\beta}_1 - \beta_1 \leq z_{1-\alpha/2}\sqrt{V}) \\ &= Pr(z_{\alpha/2}\sqrt{V} - \hat{\beta}_1 \leq -\beta_1 \leq z_{1-\alpha/2}\sqrt{V} - \hat{\beta}_1) \\ &= Pr(\hat{\beta}_1 - z_{1-\alpha/2}\sqrt{V} \leq \beta_1 \leq \hat{\beta}_1 - z_{\alpha/2}\sqrt{V}) \end{aligned}$$

Pay careful attention to the switch from lines 3 to 4. The multiplying of the inequalities by -1 switches the upper and lower thresholds.

In this case, we can use the symmetry of the normal distribution. Since the normal distribution is symmetric, $-z_{\alpha/2} = z_{1-\alpha/2}$. This gives us the symmetric CI:

$$CI_{1-\alpha} = [\hat{\beta}_1 - z_{1-\alpha/2}\sqrt{V}, \hat{\beta}_1 + z_{1-\alpha/2}\sqrt{V}]$$

However, not all distributions are symmetric. For example, can you solve the confidence interval of σ^2 , using the estimator s^2 ?

Since, the t-distribution is also symmetric, we can define the a very similar CI using \hat{V} :

$$CI_{1-\alpha} = [\hat{\beta}_1 - t_{n-k, 1-\alpha/2}\sqrt{\hat{V}}, \hat{\beta}_1 + t_{n-k, 1-\alpha/2}\sqrt{\hat{V}}]$$

where $t_{n-k, 1-\alpha/2}$ is the $1 - \alpha/2$ percentile of the t-distribution (with $n - k$ dof).

Hypothesis tests

We will consider both two-sided and one-sided hypothesis tests. The former are actually simpler than the latter, since they (typically) involve sharp null hypotheses.

Two-sided hypothesis tests

Consider the two-sided hypothesis:

$$\begin{array}{l} H_0 : \beta_1 = r \\ \text{against } H_1 : \beta_1 \neq r \end{array}$$

Here, r is just a constant (i.e. non-random scalar). For example, $r = 0$ is a simple test for whether β_1 is statistically significant.

A test requires a rejection rule that determines when you reject H_0 based on the value of the test-statistic. Consider, the T-statistic:

$$\text{T-stat} = \frac{\hat{\beta}_1 - r}{\sqrt{\hat{V}}}$$

Under $H_0 : \beta_1 = r$, implying that the T-statistic has a t-distribution. Note, it has this distribution only under H_0 . The statistic is a R.V. with infinite support. It has the continuous distribution over the real line, meaning that with a non-zero probability it can take on any value. If this is the case, any realization of the statistic (in by implication the estimator $\hat{\beta}_1$) is consistent with H_0 .

The goal is to control the probability of type 1 errors: the probability of rejecting H_0 when it is true. We do so, by choosing a significance level α which will determine the size of the test.²

Consider the rejection rule:

- Reject H_0 if $\text{T-stat} \leq t_{n-k, \alpha/2}$ or $\text{T-stat} \geq t_{n-k, 1-\alpha/2}$.

The probability of a type 1 error is:

$$\begin{aligned} Pr(\text{Reject } H_0 | \beta_1 = r, X) &= Pr(\text{T-stat} \leq t_{n-k, \alpha/2} | \beta_1 = r, X) \\ &\quad + Pr(\text{T-stat} \geq t_{n-k, 1-\alpha/2} | \beta_1 = r, X) \\ &= \alpha/2 + \alpha/2 \\ &= \alpha \end{aligned}$$

This test has size α . Moreover, given the symmetry of the t-distribution, we can write the rejection rule as:

- Reject H_0 if $|\text{T-stat}| \geq t_{n-k, 1-\alpha/2}$.

²We say that a test has size α if the probability of a type 1 error is $\leq \alpha$.

One-sided hypothesis tests

Consider the one-sided hypothesis:

$$\begin{array}{l} H_0 : \beta_1 \leq r \\ \text{against } H_1 : \beta_1 > r \end{array}$$

The null hypothesis is no longer sharp. Any value of \leq satisfies the null. It therefore becomes more difficult to think about the size of the test.

Consider the rejection rule:

- Reject H_0 if T-stat $\geq t_{n-k, 1-\alpha}$.

We reject H_0 if the value of the test statistic is much greater than the $1 - \alpha$ percentile of the t-distribution. Does this test have the right size?

$$\begin{aligned} Pr(\text{Reject } H_0 | \beta_1 \leq r, X) &= Pr(\text{T-stat} \geq t_{n-k, 1-\alpha} | \beta_1 \leq r, X) \\ &\leq Pr(\text{T-stat} \geq t_{n-k, 1-\alpha} | \beta_1 = r, X) \\ &= \alpha \end{aligned}$$

The inequality appears, because for values of $\beta_1 < r$, the probability of rejection is strictly $< \alpha$. This is not a problem, so long as for the threshold case $\beta_1 = r$, the probability is still $\leq \alpha$.

P-values

The p-value of a test corresponds to the smallest significance level (i.e. α) at which you reject H_0 . It is an incredibly useful value to compute, because it can be used as a rejection rule:

- Reject H_0 if p-value $\leq \alpha$.

This provides you with a rejection rule that is independent of the distribution of the test-statistic. The critical values of a test-statistic will depend on the distribution, but the p-value not.

In the above two-sided test, the p-value is given by

$$\text{p-value} = 2 \times (1 - T_{n-k}(|\text{T-stat}|))$$

where $T_{n-k}(\cdot)$ is the CDF of the t-distribution (with $n - k$ dof).

For the one-sided test we considered, it would be:

$$\text{p-value} = 1 - T_{n-k}(\text{T-stat})$$

While the p-value is computed using the CDF of the test-statistic, and is a value $\in [0, 1]$, **it is not a probability**. However, since it is a function of the test-statistic, it is a random variable.

One interesting feature of the p-value is that it has a uniform distribution. Consider, the probability that the p-value is less than some value ρ . We will use the one-sided test to simplify things.

$$\begin{aligned} Pr(\text{p-value} \leq \rho | X, \beta_1 = r) &= Pr(1 - T_{n-k}(\text{T-stat}) \leq \rho | X, \beta_1 = r) \\ &= Pr(T_{n-k}(\text{T-stat}) \geq 1 - \rho | X, \beta_1 = r) \\ &= Pr(\text{T-stat} \geq T_{n-k}^{-1}(1 - \rho) | X, \beta_1 = r) \\ &= 1 - T_{n-k}(T_{n-k}^{-1}(1 - \rho)) \\ &= 1 - (1 - \rho) \\ &= \rho \end{aligned}$$

Thus, the p-value has the characteristic of a uniformly distributed random variable: for $X \sim U(0, 1) \Rightarrow Pr(X \leq x) = x$.

This fact is used to evaluate p-hacking and publication bias in published research. By collecting and plotting the distribution of p-values from published research, you can test how much the distribution varies from uniform.

Statistical Power

Statistical power refers to the probability of rejecting H_0 for a given value of the unknown parameter, which need not correspond to the value under the null hypothesis. Consider the sharp null from the two-sided test: $H_0 : \beta_1 = r$. Suppose, the true value of β_1 is some other value κ .

$$Pr(\text{Reject } H_0 | \beta_1 = \kappa, X) = Pr(|\text{T-stat}| \geq t_{n-k, 1-\alpha/2} | \beta_1 = \kappa, X)$$

The T-statistic has a t-distribution only under the null hypothesis (H_0 is true). If $\beta_1 = \kappa \neq r$, then this probability is not α .

$$\begin{aligned}
Pr(\text{Reject } H_0 | \beta_1 = \kappa, X) &= Pr\left(\frac{\hat{\beta}_1 - r}{\sqrt{\hat{V}}} \leq t_{n-k, \alpha/2} \middle| \beta_1 = \kappa, X\right) \\
&\quad + Pr\left(\frac{\hat{\beta}_1 - r}{\sqrt{\hat{V}}} \geq t_{n-k, 1-\alpha/2} \middle| \beta_1 = \kappa, X\right) \\
&= Pr\left(\frac{\hat{\beta}_1 - \kappa}{\sqrt{\hat{V}}} + \frac{\kappa - r}{\sqrt{\hat{V}}} \leq t_{n-k, \alpha/2} \middle| \beta_1 = \kappa, X\right) \\
&\quad + Pr\left(\frac{\hat{\beta}_1 - \kappa}{\sqrt{\hat{V}}} + \frac{\kappa - r}{\sqrt{\hat{V}}} \geq t_{n-k, 1-\alpha/2} \middle| \beta_1 = \kappa, X\right) \\
&= Pr\left(\frac{\hat{\beta}_1 - \kappa}{\sqrt{\hat{V}}} \leq t_{n-k, \alpha/2} - \frac{\kappa - r}{\sqrt{\hat{V}}} \middle| \beta_1 = \kappa, X\right) \\
&\quad + Pr\left(\frac{\hat{\beta}_1 - \kappa}{\sqrt{\hat{V}}} \geq t_{n-k, 1-\alpha/2} - \frac{\kappa - r}{\sqrt{\hat{V}}} \middle| \beta_1 = \kappa, X\right)
\end{aligned}$$

Under the condition that $\beta_1 = \kappa$,

$$\frac{\hat{\beta}_1 - \kappa}{\sqrt{\hat{V}}} \sim T_{n-k}$$

$$\begin{aligned}
Pr(\text{Reject } H_0 | \beta_1 = \kappa, X) &= Pr\left(T \leq t_{n-k, \alpha/2} - \frac{\kappa - r}{\sqrt{\hat{V}}} \middle| \beta_1 = \kappa, X\right) \\
&\quad + Pr\left(T \geq t_{n-k, 1-\alpha/2} - \frac{\kappa - r}{\sqrt{\hat{V}}} \middle| \beta_1 = \kappa, X\right) \\
&\geq \alpha
\end{aligned}$$

We know that this probability is at least as large as α , the probability of a type 1 error. As a function of κ , this probability will tend towards 1, as κ moves further from the null hypothesis. For a two-sided test, it does so symmetrically, creating a bell-shaped function in the case of the normal and t-distributions. Note, as $|\kappa - r|$ increases, the one probability increases, while the other decreases. Regardless, the power increases because the one probability increases faster than the other decreases. This is a result of the bell-shaped t-distribution.

For a one-sided test, the power function will asymptote to 1 only on the side of rejection. We say that a test is uniformly more powerful, if has greater statistical power for all possible values of the parameter.

When comparing tests that have the same distribution, the difference in power will arise from the variance of the estimator \hat{V} . A more efficient estimator, will yield a more power test.