Regression #5: Confidence Intervals and Hypothesis Testing (Part 1)

Econ 671

Purdue University

Introduction

- What is a confidence interval?
- To fix ideas, suppose you sit down at STATA and run a regression. The standard output consists of a set of point estimates, standard errors, t-statistics, F-statistics, and confidence intervals.
- Suppose that, for the parameter you care about, the 95 % confidence interval is reported as:

  $$[.3, .7]$$

  What does this mean? How would you interpret this result?

Introduction

I don't know whether or not you got this "right," but what we commonly see is an interpretation like the following one:

The parameter that I am most interested in, β_j, lies in the reported interval $[.3, .7]$ *with 95% probability.*

This is the *wrong* interpretation since β_j is fixed, and the above attributes randomness to this population parameter. It will either fall within a given interval, or it will not.

Introduction

The correct interpretation is important, in my view, and somewhat non-intuitive:

Suppose we had the luxury of getting data sets over and over again from the population of interest, each with sample size n.

Each time we conduct this (thought) experiment, we apply our estimator to the new sample of data, and each time, construct a new "confidence interval."

Upon repeated sampling in this way, the collection of the intervals that I obtain from this process will contain the "true" β_j *95 percent of the time.*

Though it is indeed seductive to ascribe some type of probability of content to any particular realized interval, this is not the correct interpretation.

Some Preliminaries

- In the following few slides, we will provide some derivations and establish results we have already assumed.
- The purpose of this lecture is to rigorously derive the distribution of what we commonly use as a "test statistic." We do so conditionally on X and under normality.
- Later lectures will consider generalizations of this basic result.

Some Preliminaries

Suppose

$$x \sim \mathcal{N}(\mu, \Sigma).$$

Theorem (**Linear combinations of normals are normal)**

Let $y = \theta + Hx$ *for a* $k \times k$ *non-singular matrix* H *and a* $k \times 1$ *vector* x*. Then,*

$$y \sim \mathcal{N}\left(\theta + H\mu, H\Sigma H'\right).$$

Since H is non-singular we can write:

$$x = H^{-1}(y - \theta)$$

whence the Jacobian of the transformation from x to y is:

$$|det(H^{-1})|.$$

Some Preliminaries

Thus,

$$p(y) = (2\pi)^{-k/2} |det(H^{-1})| det(\Sigma^{-1})^{1/2} \exp\left(-\frac{1}{2}(H^{-1}(y-\theta)-\mu)'\Sigma^{-1}(H^{-1}(y-\theta)-\mu)\right)$$

Proof.

First, consider the quadratic form in the exponential kernel:

$$\begin{aligned}
\left[H^{-1}(y-\theta)-\mu\right]' \quad \Sigma^{-1} \quad & \left[H^{-1}(y-\theta)-\mu\right] \\
= \quad & \left[H^{-1}(y-\theta-H\mu)\right]' \Sigma^{-1} \left[H^{-1}(y-\theta-H\mu)\right] \\
= \quad & (y-\theta-H\mu)'(H')^{-1}\Sigma^{-1}H^{-1}(y-\theta-H\mu) \\
= \quad & \left[y-(\theta+H\mu)\right]' \left(H\Sigma H'\right)^{-1} \left[y-(\theta+H\mu)\right]
\end{aligned}$$

□

Some Preliminaries

Proof.

It remains to consider the normalizing constant. To this end, note:

$$
\begin{aligned}
|det(H^{-1})|[det(\Sigma^{-1})]^{1/2} &= \left[\left(det(H^{-1})\right)^2 det(\Sigma^{-1})\right]^{1/2} \\
&= \left[det(H^{-1})det[(H')^{-1}]det(\Sigma^{-1})\right]^{1/2} \\
&= \left[det(H)det(\Sigma)det(H')\right]^{-1/2} \\
&= \left[det(H\Sigma H')\right]^{-1/2}
\end{aligned}
$$

□

Putting these two pieces together shows that

$$y \sim \mathcal{N}\left(H\mu + \theta, H\Sigma H'\right)$$

as desired.

Some Preliminaries

We now prove that a particular quadratic form of normal random variables follows a χ^2 distribution:

Theorem

Suppose $x \sim N(\mu, \Sigma)$ where x is $k \times 1$. Then,

$$(x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2_k.$$

We will prove this using moment generating functions, basically as a means of review. (When these exist, they are unique and uniquely determine the distribution).

Some Preliminaries

Let

$$y = (x - \mu)'\Sigma^{-1}(x - \mu)$$

Note:

$$\begin{aligned}
E[\exp(ty)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(ty)\phi(x; \mu, \Sigma) dx_1 \cdots dx_k \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(1-2t)y\right) dx_1 \cdots dx_k \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'[\Sigma(1-2t)^{-1}]^{-1}(x-\mu)\right) \\
&= (1-2t)^{-k/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-k/2} |\Sigma(1-2t)^{-1}|^{-1/2} \\
&\quad \times \exp\left(-\frac{1}{2}(x-\mu)'[\Sigma(1-2t)^{-1}]^{-1}(x-\mu)\right) dx_1 \cdots dx_k \\
&= (1-2t)^{-k/2}
\end{aligned}$$

This is recognized as the m.g.f. of a chi-square random variable with k degrees of freedom.

Derivation of Test Statistic

To begin, we will obtain and derive a distribution of a test static under the assumptions that:

$$\epsilon | X \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2 I_n)$$

and, moreover, we will also condition on X.

The normality assumption yields the sampling distribution of $\hat{\beta}$:

-

Derivation of Test Statistic

Using our previous theorem regarding quadratic forms of multivariate normals, we can write:

-

What keeps you from applying this as a test statistic?

Some Preliminaries

Before getting into the details, we first digress for a while:

Theorem (Diagonalization of a symmetric matrix)

Let A be a symmetric $n \times n$ matrix.
Then $\exists$ a nonsingular matrix S with $S'S = SS' = I_n$ and diagonal Λ such that:

$$S'AS = \Lambda.$$

A consequence of this result is the representation:

$$SS'ASS' = A = S\Lambda S'.$$

Some Preliminaries

Theorem (**The rank of a symmetric matrix equals the number of non-zero eigenvalues)**

Put a little differently, for a symmetric $A = S\Lambda S'$,

$$rank(A) = rank(\Lambda).$$

The diagonal elements of Λ are the *eigenvalues* of A and S is an orthonormal matrix.

Finally, note that since Λ is diagonal, its rank is simply the number of non-zero elements along the diagonal (i.e., non-zero eigenvalues).

Some Preliminaries

Proof.

Note that, since S is non-singular,

-

The last two lines follow since the rank of a product can not exceed the rank of any of the constituent matrices in the product. It follows that

-

□

Some Preliminaries

Now, let us consider a further result for the the specific case of a symmetric, idempotent matrix M.

Theorem (The rank of a symmetric idempotent matrix equals its trace)

$$rank(M) = tr(M).$$

Proof.

We proceed in two steps. First we show that the eigenvalues (elements of Λ) must be either zero or one.

To see this, note

-

For this to be true, the diagonal elements of Λ must either be zero or one. □

Some Preliminaries

The second part of our proof notes that

-

and this trace is simply the number of unit elements on the diagonal of Λ. Our previous theorem showed that the rank of M was the rank of Λ, which is also the number of unit elements on the diagonal of Λ. Thus, when M is symmetric and idempotent,

-

Some Preliminaries

We are now ready to prove our main theorem:

Theorem

Suppose a $n \times 1$ vector x has a standard multivariate normal distribution:

$$x \sim \mathcal{N}(0, I_n)$$

and consider the quadratic form:

-

where M is symmetric and idempotent. It follows that

-

Some Preliminaries

Proof.

Since M is symmetric, employ the decomposition:

-

It follows that

-

and likewise for $S'x$. Let $z = S'x$ so that

-

This quadratic form is the sum of squares of J independent standard normal random variables, where $J = \text{tr}(M)$ is the number of unit elements in Λ. Thus, the proof is completed. □

Some Preliminaries

We can now use this result to characterize the sampling distribution of $\hat{\sigma}^2$:

-

This last equation is a quadratic form of a standard normal vector, and M is idempotent. Our previous theorem thus establishes:

-

since $\text{tr}(M) = n - k$.

Derivation of Test Statistic

OK. so now let's go way back and consider the following statistic:

-

Our previous results show that this is the ratio of two chi-square random variables divided by their respective degrees of freedom. If these random variables are *independent*, the ratio will have a $F_{k,(n-k)}$ distribution. Canceling terms, then, we suspect:

-

Derivation of Test Static

The fact that these random variables are independent is suggested by the following reasoning:

Let

-

and consider the distributions of

-

Both of these quantities are (conditionally) normally distributed.
Moreover, letting $P_x = X(X'X)^{-1}X'$:

-

Derivation of Test Static

The following points complete the proof:

- Zero covariance between *normal* random variables implies the stronger condition of *statistical independence* between those variables.
- $\hat{\sigma}^2$ depends only on $y - \hat{\mu} = \hat{\epsilon}$.
- The fitted residuals tell us nothing about the regression coefficients β. To this end, note:
- Thus, independence between $\hat{\mu}$ and $y - \hat{\mu}$ will imply independence between $\hat{\sigma}^2$ and $\hat{\beta}$.

Derivation of Test Static

To conclude, we have established that conditionally on X, and under normality,

-

a result that we will employ for testing hypotheses regarding the elements of β.