**LECTURE 16**
**INTRODUCTION TO STATISTICS: INFERENCE**

Recall the Normal location-scale model discussed in Lecture 15. The econometrician is interested in learning about the mean $\mu$ of the distribution $N(\mu, \sigma^2)$ using $n$ iid draws from that distribution: $Y_1, \ldots, Y_n$. We proposed to estimate $\mu$ using the average of $Y_i$'s:

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

We showed that the estimator $\hat{\mu}$ is a random variable with the following properties:

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

One of the implications of the above result is $\hat{\mu}$ will be different from the true parameter $\mu$ with probability equal to one:

$$P\left(\hat{\mu} = \mu\right) = 0.$$

Thus, with finite amount of data, one can never learn the true population value of the parameter. Nevertheless, we expect $\hat{\mu}$ to be close to/informative about the true value $\mu$, especially when $n$ is large. At the inference stage, the econometrician must decide whether certain values of $\mu$ are supported or not by the observed data.

# 1 Hypotheses testing

**Definition 1.** A statistical hypothesis is an assertion about the value of an unknown parameter.

For example, in the Normal location-scale model, hypotheses about $\mu$ can be:

$$\begin{aligned}
\mu &= 0, \\
\mu &= 100, \\
\mu &> 0, \\
\mu &\geq 0, \\
\mu &\neq 0,
\end{aligned}$$

and etc. In a typical hypotheses testing problem, we will see two competing hypotheses, denoted $H_0$ and $H_1$. In the simplest case, $H_0$ and $H_1$ will take the following form:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1,$$

where $\mu_0$ and $\mu_1$ are two numbers such that $\mu_0 \neq \mu_1$. For example,

$$H_0 : \mu = 0, \quad H_1 : \mu = 1.$$

In that example, the econometrician must decide whether the data $Y_1, \ldots, Y_n$ is drawn from the normal distribution with mean zero or mean one. Note that this problem is much simpler than the estimation problem discussed in the previous lecture: the econometrician does not have to try and guess the true value of $\mu$ (which involves choosing among infinitely many possibilities), but rather decide between only two values.

**Definition 2.** We say that a hypothesis is *simple* if it specifies a unique value of the unknown parameter. If a hypothesis specifies more than one value for the parameter, it is said to be *composite*.

We often encounter situations where one or both hypotheses are composite:

$$H_0 : \mu = 0, \quad H_1 : \mu \neq 0,$$
$$H_0 : \mu \leq 0, \quad H_1 : \mu > 0,$$

and etc. Note that $H_0$ and $H_1$ must be disjoint:

$$H_0 \cap H_1 = \emptyset,$$

i.e. they cannot be both true simultaneously. At the same time, we assume that either $H_0$ or $H_1$ must be true, i.e. the true value $\mu$ satisfies

$$\mu \in (H_0 \cup H_1).$$

**Definition 3.** The union of $H_0$ and $H_1$ is called the *maintained* hypothesis.

The roles of $H_0$ and $H_1$ are not arbitrary. It is agreed that the null hypothesis ($H_0$) should be treated as true unless the data provides sufficient evidence against it. Hence, $H_1$ is often referred to as the *alternative* hypothesis. The testing problem is viewed as testing of $H_0$ *against* $H_1$. It is also agreed that the assertion that the econometrician is interested in showing as true must be stated as $H_1$. Therefore, in such a setup, the econometrician must carry the burden of proof.

**Definition 4.** Statistical *test* is a decision rule for choosing between $H_0$ and $H_1$.

**Definition 5.** Test *statistic* is any function of the data $Y_1, \ldots, Y_n$ that is used in the definition of a test.

Typically, a test statistic measures the distance between the data and the null hypothesis. When this distance is large, one would reject the null hypothesis in favor of the alternative. The range of possible values for the test statistic is divided into two regions: the *acceptance* region and the *rejection* (or *critical*) region. The acceptance region corresponds to a small distances between the test statistic and the null hypothesis; and the rejection region corresponds to large distances between the test statistic and the null. A statistical test is formulated as a *rejection rule*: Reject $H_0$ when the test statistic is in the rejection region.

Thus, constructing a statistical test comes down to choosing a test statistic and selecting a decision rule. Note that, when $H_0$ is true, it is possible for the distance between the statistic and $H_0$ to be large. This can happen due to randomness of the data. For example, when data are drawn from $N(0, \sigma^2)$, it is possible for finitely many draws $Y_i$'s to be "far" from zero, even though such an event would occur only with a small probability. Thus, a test statistic can end up in the rejection region even though the null hypothesis is true. Similarly, when $H_1$ is true, a test statistic may end up in the acceptance region. Thus, our main concern here is making a wrong decision and evaluating the probability of making a wrong decision.

When there are two possible states of the world ($H_0$ is true or $H_1$ is true), and the econometrician must choose between $H_0$ and $H_1$, there are four possible outcomes:

| truth \ choice | $H_0$ | $H_1$ |
|---|---|---|
| $H_0$ | ✓ | Type I error |
| $H_1$ | Type II error | ✓ |

**Definition 6. (Type I and II errors)** Rejecting $H_0$ when it is true is called Type I error. Accepting $H_0$ when it is false is called Type II error.

Type I error our first and foremost concern. This is because the econometrician has to carry the burden of proof: the hypothesis he is interested in is stated as $H_1$, and the econometrician is supposed to find strong evidence in the data against $H_0$. Thus, our first concern is with rejecting $H_0$ when in fact it is true. Hence, Type I error can be viewed as a *false discovery.*

Ideally, we would like to have the probabilities of both errors, Type I and II, to be as small as possible. Unfortunately, typically there is a tradeoff between the two. Once it has been decided what test statistic to use, a test and the probabilities of Type I and II errors would depend only on the definition of the rejection region. To reduce the probability of Type I error, one would have to shrink the rejection region. While this would reduce the probability of rejecting $H_0$ when it is true (and the probability of Type I error), it would also reduce the probability of rejecting $H_0$ when it is false and, therefore, will increase the probability of Type II error.

**Definition 7. (Size and power)** The probability of rejecting $H_0$ when it is true (the probability of Type I error) is called the *size* of a test. The probability of rejecting $H_0$ when it is false (one minus the probability of Type II error) is called the *power* of a test.

Since our first concern is with false discoveries, we therefore define the *validity* of a test through size control.

**Definition 8. (Validity of a test/size control)** Let $\alpha \in (0, 1)$ be a pre-specified *significance level.* We say that a test is valid if the probability of rejecting $H_0$ when it is true (the probability of Type I error) does not exceed $\alpha$.

The significance level $\alpha$ is selected to be a small number (close to zero): the commonly used values are: 0.05, 0.01, and 0.10. For example, when $\alpha = 0.05$, we only allow tests that have the probability of type one error under 5%. Tests that reject $H_0$ when it is true with a higher probability would be invalid.

Given two valid tests, we would say that one of them is *more powerful* if it rejects $H_0$ with a higher probability when it is false. Such a test would be more desirable as it has a smaller probability of Type II error.

Below, we illustrate the discussed concepts using the Normal model.

## 2 Normal location model

Suppose that $Y_1, \dots, Y_n$ are iid $N(\mu, \sigma^2)$. Assume for simplicity that $\sigma^2$ is *known*. Thus, the only unknown parameter is the mean $\mu$. We are interested in testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

Such an alternative hypothesis (and the corresponding test) is called *two-sided*, since $H_1$ allows deviations from $H_0$ in either direction. Note that $\mu_0$ is a known number specified by the econometrician.

The information about $\mu$ in the data is summarized by the estimator

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Note that, since both $\sigma^2$ and $n$ are known in this model, the variance of the distribution of $\hat{\mu}$ is known. We need to decide whether the data supports $\mu = \mu_0$ or $\mu \neq \mu_0$.

We define the test statistic as the distance between the null hypothesis (or $\mu_0$) and the data (or $\hat{\mu}$):

$$|S(\mu_0)| = \left|\frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}\right|,$$

where

$$S(\mu_0) = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}.$$

The distribution of $\hat{\mu}$ is centered around the true value $\mu$, and the estimator $\hat{\mu}$ is expected to be in its neighborhood. If in our sample $\hat{\mu}$ is far from $\mu_0$, it would be plausible to argue that $\mu_0 \neq \mu$. Thus, a large distance between $\hat{\mu}$ and $\mu_0$ can be taken as evidence against $H_0$ and in favor of $H_1$. It is important that we do not simply use the absolute distance between $\hat{\mu}$ and $\mu_0$, but re-scale it using the standard deviation of the distribution of $\hat{\mu}$, which is given by $\sigma/\sqrt{n}$.[1]

We should reject $H_0$ when the distance between $\mu_0$ and $\hat{\mu}$ is large:

$$\text{Reject } H_0 \text{ when } |S(\mu_0)| = \left|\frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}\right| > c,$$

where $c > 0$ is some constant. This constant $c$ is called the *critical* value. Thus, the corresponding *rejection* or *critical* region for $S(\mu_0)$ is given by

$$\text{Rejection region} = (-\infty, -c) \cup (c, \infty),$$

and the acceptance region is given by

$$\text{Acceptance region} = [-c, c].$$

[1]"Large distance" means different things in distributions with different variances. When the variance is equal to 1, the absolute distance of 2 from the mean can be considered large: the probability of drawing a value that deviates from the mean by 2 or more is under 5% (approximately). At the same time, if the variance is equal to 100, the probability of drawing a such a value is over 84%. Hence, in statistics "large" must be defined relatively to the variance/standard deviation of the distribution.

The constant $c$ has to be chosen so that the size of the test is controlled, i.e. the probability of Type I error does not exceed some pre-specified value $\alpha$. The probability of Type I error is given by the probability of $S(\mu_0) \in$ Rejection region. Hence, $c$ must satisfy the following condition: when $\mu = \mu_0$

$$P\left(\left|\frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}\right| > c \,|\mu = \mu_0\right) = P\left(\left|\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}\right| > c\right) = \alpha.$$

Since $\hat{\mu} \sim N(\mu, \sigma^2/n)$, it follows by the properties of the normal distribution that

$$\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Denoting

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

we find that the critical value $c$ is determined by

$$P\left(|Z| > c\right) = \alpha,$$

which must be solved for $c$ in terms of $\alpha$. The solution is described in the following lemma.

**Lemma 9.** *Let $Z \sim N(0,1)$. The constant $c$ in $P(|Z| > c) = \alpha$ satisfies*

$$c = z_{1-\alpha/2},$$

*where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of $N(0,1)$ distribution.*

*Remark.* When $\alpha = 0.05$, the critical value $c$ is given by $z_{1-0.05/2} = z_{0.975} \approx 1.96$. When $\alpha = 0.01$, the corresponding critical value is $z_{1-0.01/2} = z_{0.995} \approx 2.58$. When $\alpha = 0.10$, the corresponding critical value is $z_{1-0.1/2} = z_{0.95} \approx 1.64$.

*Proof.* Recall that from the definition of quantiles (see Theorem 2 in Lecture 10),

$$P\left(Z \leq z_{1-\alpha/2}\right) = 1 - \alpha/2,$$

and

$$P\left(Z > z_{1-\alpha/2}\right) = \alpha/2.$$

Thus,

$$\begin{aligned} P\left(|Z| > z_{1-\alpha/2}\right) &= P\left(Z > z_{1-\alpha/2}\right) + P\left(Z < -z_{1-\alpha/2}\right) \\ &= \alpha/2 + \alpha/2 \\ &= \alpha. \end{aligned}$$

□

To summarize, when $\sigma$ is known, our size-$\alpha$ test for $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ depends on the test statistic

$$S(\mu_0) = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}.$$

The test is given by

$$\text{Reject } H_0 \text{ when } |S(\mu_0)| > z_{1-\alpha/2}.$$

Lemma 9 establishes, that the probability of Type I error for our test is exactly $\alpha$, and therefore the test is a valid size $\alpha$ test.

## 3 Power

It is important also to evaluate the probability of Type II error or the power of the test, where the latter is given by the probability of rejecting $H_0$ when it is false. For that purpose, we need to know the distribution of our statistic $S(\mu_0)$ under $H_1 : \mu \neq \mu_0$. Write

$$\begin{aligned} S(\mu_0) &= \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \\ &= Z + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \\ &\sim N\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}}, 1\right). \end{aligned}$$

While under $H_0 : \mu = \mu_0$ the distribution of $S(\mu_0)$ is centered around zero, under alternative it is shifted by

$$\frac{\mu - \mu_0}{\sigma/\sqrt{n}}.$$

Note that the variance of $S(\mu_0)$ remains the same under $H_0$ and $H_1$. The shifting of the distribution of $S(\mu_0)$ puts more probability mass into the rejection region, and as a result the probability of rejection would be greater than the null rejection probability $\alpha$:

$$\begin{aligned} P\left(|S(\mu_0)| > z_{1-\alpha/2}\right) &= P\left(S(\mu_0) > z_{1-\alpha/2}\right) + P\left(S(\mu_0) < -z_{1-\alpha/2}\right) \\ &= P\left(Z + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha/2}\right) + P\left(Z + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha/2}\right) \\ &= P\left(Z > z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) + P\left(Z < -z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right). \end{aligned}$$

Note that the only unknown quantity in the above expression is the true value of the parameter $\mu$. Thus, the power of the test is a function of $\mu$: we can compute it after substituting values for $\mu$. Let

$\pi(\mu)$ denote the power as a function of $\mu$, i.e.

$$\begin{aligned}
\pi(\mu) &= P\left(|S(\mu_0)| > z_{1-\alpha/2}\right) \\
&= P\left(Z > z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) + P\left(Z < -z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right).
\end{aligned}$$

**Example 10.** Suppose that $\sigma = 1$, $n = 100$, $\mu_0 = 0$, and $\alpha = 0.05$, so that $z_{1-\alpha/2} = 1.96$. In that case,

$$\pi(\mu) = P\left(Z > 1.96 - 10\mu\right) + P\left(Z < -1.96 - 10\mu\right).$$

We obtain:

$$\pi(\mu) = \begin{cases} 0.9988, & \mu = -0.5 \\ 0.516, & \mu = -0.2 \\ 0.1701, & \mu = -0.1 \\ 0.05, & \mu = 0 \\ 0.1701, & \mu = 0.1 \\ 0.516, & \mu = 0.2 \\ 0.9988, & \mu = 0.5 \end{cases}$$

Note that the power of the test is increasing in the distance between $\mu$ and $\mu_0$: $H_0$ is more likely to be rejected when $\mu$ is further away from $\mu_0$. Note also that the power is symmetric around $\mu_0 = 0$. This is a feature of two-sided tests in the case of symmetric distributions. Lastly, a test is more powerful when there is more data: a large value of $n$ would correspond to the shift

$$\frac{\mu - \mu_0}{\sigma/\sqrt{n}}$$

of a larger magnitude. A bigger portion of the distribution would be in the rejection region when $\mu \neq \mu_0$,which would correspond to a higher probability of rejection. For example, suppose that $n = 400$. In that case,

$$\pi(\mu) = P\left(Z > 1.96 - 20\mu\right) + P\left(Z < -1.96 - 20\mu\right),$$

and

$$\pi(\mu) = \begin{cases} 0.9999, & \mu = -0.5 \\ 0.9793, & \mu = -0.2 \\ 0.516, & \mu = -0.1 \\ 0.05, & \mu = 0 \\ 0.516, & \mu = 0.1 \\ 0.9793, & \mu = 0.2 \\ 0.9999, & \mu = 0.5 \end{cases}$$

# 4 $p$-values

The rejection region of the test described in the previous section depends on the critical value $z_{1-\alpha/2}$, which in turn depends on the *significance level* $\alpha$. Larger values of $\alpha$ correspond to a higher probability of Type I error, and therefore to larger rejection regions. Therefore, it is easier to reject $H_0$ in the case of a test with a bigger significance level $\alpha$: it is possible that one rejects $H_0$ in the case of a test with significance level $\alpha_1$, but then accepts $H_0$ in the case of a test with significance level $\alpha_2 < \alpha_1$.

**Definition 11.** Given the value of a test statistic, the $p$-value is the *smallest* significance level that allows to reject $H_0$.

It follows from the definition of $p$-values, that a statistical test with significance level $\alpha$ can be alternatively stated as

$$\text{Reject } H_0 \text{ if } p\text{-value}<\alpha.$$

For the two-sided test discussed in the previous section the $p$-value can be computed as follows. The test is given by

$$\text{Reject } H_0 \text{ if } |S(\mu_0)| > c$$

or

$$\text{Reject } H_0 \text{ if } S(\mu_0) > c \text{ or } S(\mu_0) < -c.$$

Our decision would switch from rejection to acceptance exactly at $c = S(\mu_0)$. Hence, we need to find the tail probabilities corresponding to the probability mass to the right of $|S(\mu_0)|$ and to the left of $-|S(\mu_0)|$. Let $\Phi(z)$ denote the standard normal CDF (recall that the null distribution of $S(\mu_0)$ is $N(0,1)$). The probability mass to the right of $|S(\mu_0)|$ is given by

$$1 - \Phi(|S(\mu_0)|).$$

The probability mass to the left of $-|S(\mu_0)|$ is given by

$$\Phi(-|S(\mu_0)|) = 1 - \Phi(|S(\mu_0)|),$$

where the equality holds due to the symmetry of the standard normal distribution around zero. Hence the $p$-value is given by

$$\begin{aligned} p\text{-value} &= 1 - \Phi(|S(\mu_0)|) + \Phi(-|S(\mu_0)|) \\ &= 2\left(1 - \Phi(|S(\mu_0)|)\right). \end{aligned} \tag{1}$$

*Remark.* Note that while $p$-values are between zero and one, they are not probabilities. From (1) it is clear that the $p$-value is a transformation of the test statistic, which is a random variable. Hence, $p$-values are random variables. Moreover, $p$-values are test statistics transformed to the zero-one scale. The corresponding critical value is simply $\alpha$.

# 5 Confidence intervals

When testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, the null hypothesis will be rejected for certain values of $\mu_0$ and accepted for other values of $\mu_0$. Rejected $\mu_0$'s are unlikely to be the true value of the parameter $\mu$as they are inconsistent with the data. On the other hand, accepted $\mu_0$'s are plausible to be the true value of the parameter. Collecting all *non-rejected* $\mu_0$'s would give a set of values (for the true parameter $\mu$) that are consistent with the data. This set is called *the confidence set* or *confidence interval* (as we will see later, in the case of our two-sided test in Section 2), the confidence set is an interval.

**Definition 12. (Confidence Interval)** Confidence interval with the *confidence level* $1-\alpha$ *is the set of all values* $\mu_0$ *that where not rejected by a test of* $H_0 : \mu = \mu_0$ *against* $H_1 : \mu \neq \mu_0$ *by a size* $\alpha$ *test.*

For the normal location model in Section 2, and the two-sided test which rejects $H_0$ when $|S(\mu_0)| > z_{1-\alpha/2}$, the confidence interval is given by

$$\begin{aligned}
CI_{1-\alpha} &= \left\{\mu_0 : |S(\mu_0)| \leq z_{1-\alpha/2}\right\} \\
&= \left\{\mu_0 : -z_{1-\alpha/2} \leq S(\mu_0) \leq z_{1-\alpha/2}\right\} \\
&= \left\{\mu_0 : -z_{1-\alpha/2} \leq \frac{\hat{\mu}-\mu_0}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right\} \\
&= \left\{\mu_0 : \hat{\mu} - z_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \hat{\mu} + z_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right\} \\
&= \left[\hat{\mu} - z_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right].
\end{aligned}$$

A confidence interval has a very important feature/interpretation: it is a random interval that includes the true value $\mu$ with the probability $1-\alpha$.

**Theorem 13.** *Suppose that* $\hat{\mu} \sim N(\mu, \sigma^2/n)$. *Then,*

$$P\left(\mu \in CI_{1-\alpha}\right) = 1 - \alpha.$$

*Proof.* From the definition of the confidence interval,

$$\begin{aligned}
P\left(\mu \in CI_{1-\alpha}\right) &= P\left(|S(\mu)| \leq z_{1-\alpha/2}\right) \\
&= 1 - P\left(|S(\mu)| > z_{1-\alpha/2}\right) \\
&= 1 - \alpha.
\end{aligned}$$

The equality in the last line holds because

$$S(\mu) = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} = Z,$$

and by Lemma 9. $\square$