School of Oriental and African Studies **September 2013**
Department of Economics

Preliminary Statistics
Lecture 5: Hypothesis Testing (Outline)

Gujarati D. *Basic Econometrics*, Appendix A.8
Barrow M. *Statistics for Economics, Accounting and Business Studies* Ch. 5

1 Introduction

In testing we want to make inferences about the unknown population parameters in the form of tests of some hypotheses about the value of the population parameters. The elements of a statistical test are:

- Null Hypothesis, $H_0 : \theta = \theta_0$
- Alternative Hypothesis, H_1 :
  - Two-sided alternative $\theta \neq \theta_0$
  - One-sided alternative $\theta < \theta_0$ or $\theta > \theta_0$
- Test Statistic, a function of the sample measurements upon which the statistical decision will be based
- Decision Rule, rule which specifies for which values of the test statistic the null hypothesis will be rejected and for which it will not

Statistical testing is analogous to a criminal trial. There is a null hypothesis and an alternative hypothesis (innocent vs guilty), a test statistic (evidence against/in favour of the initial hypothesis), and a decision rule (the jury rejects the null if the evidence presented indicates the defendant guilty 'beyond reasonable doubt'). Similarly, in hypothesis testing H_0 is rejected if the evidence from the sample seems inconsistent with the hypothesis made, that is if H_0 has a very low probability of occurring if it were true.

The problem is, we can never know with certainty which of the hypotheses is true, hence two mistakes can happen. Type I error: reject H_0, though it is true (convict an innocent person) or Type II error: accept H_0 though it is false (set a guilty person free).

Table 1: Hypothesis Testing: Decision Rule

	H_0 true	H_0 false
Accept H_0	✓	Type II error
Reject H_0	Type I error	✓

We can avoid Type I error completely by always accepting the null hypothesis, but we would be making a lot of type II errors, and vice versa. Statistical tests design the test procedure so that there is a fixed risk of Type I error. This probability is known as the significance level and is usually fixed at 5%.

2 Procedure of Testing

Suppose we have a random variable Y with mean μ and variance σ^2. We want to test the hypothesis that the true mean takes a particular value μ_0, using a random sample $Y_i, i = 1, 2, ...n$.

1. State the Null Hypothesis $\mathrm{H}_0 : \mu = \mu_0$
2. State the Alternative (2-sides) Hypothesis $\mathrm{H}_1 : \mu \neq \mu_0$
3. Construct the Test Statistic under the null (i.e. if H_0 were true)

$$\tau = \frac{\hat{\mu} - \mu_0}{SE(\hat{\mu})}$$

4. Find how the Test Statistic is distributed under H_0
   (a) For large samples ($n > 30$), the test statistic is standard normally distributed
   $$\tau \sim SN(0, 1)$$
   (b) For small samples ($n < 30$) and if the true variance (σ^2) is unknown, the test statistic is distributed as a student-t with $n - 1$ degrees of freedom
   $$\tau \sim t_{n-1}$$
5. Find the critical values in the tails of the distribution of the test statistic that give (conventionally) the 95% of the distribution. That is, specify the critical values so the probability that $\hat{\mu}$ being outside the critical values is small, typically $\alpha = 5\%$. This is the significance level of the test. Other common values for α include 0.1 (10%) and 0.01 (1%).
   - For a two-sided test ($\mu \neq \mu_0$), the level of significance is equally split between the two tails in order to find the critical values. If $\alpha = 0.05$, the critical values would be found so that 0.025 is left in each side of the distribution
   - For a one-sided test ($\mu > \mu_0 \,\text{or}\, \mu < \mu_0$), the entire level of significance will be located in one tail (0.05).
6. Apply the decision rule: if the Test statistic value (in absolute value) is greater than the critical value, reject H_0. That is, if the test statistic is greater than the critical value or it is less than minus the critical value, we reject the null. Otherwise, we do not.

   An alternative method of applying the decision rule is to quote the probability of obtaining the test statistic assuming the null hypothesis is true. This is known as the p-value approach. It has the advantage of avoiding arbitrary selection of a cut-off point. In the same lines as with the critical values, if we are testing a one-side test (e.g. $\mathrm{H}_1 : \mu > \mu_0$) we find the probability of the test statistic being greater than the standardised value and compare it with 0.05. If we are doing a 2-sided test, we find the probability, multiply it with 2 and then compare it with the significance level (5%).

Note that statistical significance and substantive significance can be very different. An effect may be very small of no importance, but statistically very significant, because we have a very large sample and a small standard error. Alternatively, an effect may be large, but not statistically significant because we have a small sample and it is imprecisely estimated. Statistical significance asks: 'could the difference have arisen by chance in a sample of this size?' not 'is the difference important?'

3 Type I and Type II errors

As we said earlier, in statistical test we fix the probability of making a type one error to the significance level, α (conventionally to 5%).

$$\text{Prob}\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = (1-\alpha)$$
$$\text{Prob}\left(-z_{2.5} \leq Z \leq z_{2.5}\right) = 1 - 5\% = 95\%$$
$$\text{Prob}\left(-1.96 \leq \frac{\hat{\theta}-\theta_0}{SE(\hat{\theta})} \leq 1.96\right) = 95\% \text{ if } \mathrm{H}_0 \text{ is true}$$

So, the way we actually apply the decision rule, says that if the H_0 is true, and we find a Test Statistic value in the extreme 5% (2.5% from both sides) of the distribution of the test statistic we reject the H_0 (though it is true), which is the probability of making a type I error.

The power of a test is the probability of rejecting a false hypothesis. If we denote by β the probability of committing a Type II error, then the power of the test is $(1-\beta)$, the probability of not committing a Type II error, i.e. probability of rejecting a false null hypothesis. This is one of the correct decisions identified before. Essentially, the power of a test shows the ability of a test to detect an effect, if it actually exists. Unfortunately, decreasing the probability of error type, increases the probability of the other. Only by increasing the sample size can the probability of one error be improved without an adverse effect on the other error.

4 Confidence Interval Approach to Hypothesis Testing

There is also the confidence interval approach to hypothesis testing. According to this approach we construct a confidence interval around the point estimate, and if the hypothesised value falls outside of the interval, we reject the null hypothesis. The values within the confidence interval are the set of all acceptable hypotheses. The confidence coefficient $(1-\alpha)$ corresponds to the probability of accepting the null hypothesis while being true.

5 An Example - Basketball Players Height

5.1 Reminder

Assume the random variable 'height of basketball players', Y, is normally distributed with mean μ and variance σ^2.

$$Y \sim N(\mu, \sigma^2)$$

We have a random sample of n basketball players' height, $Y_i, i = 1, 2...n$ drawn from the population. To estimate the true parameters (μ, σ^2) we use the sample

mean ($\hat{\mu}$) as the estimator for μ and the sample variance (s^2) as the unbiased estimator for σ^2.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{n}, \qquad s^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{\mu})^2}{n-1}$$

Since Y is normally distributed, $\hat{\mu}$ will be also normally distributed (because $\hat{\mu}$ is a linear function of Y). Nonetheless, even if Y were not normally distributed, then according to the Central Limit Theorem $\hat{\mu}$ will be normal as the sample size increases. Hence, $\hat{\mu} \sim N(\mu, \sigma^2/n)$. An estimate for the standard error of the sample mean is provided by, $\text{SE}(\hat{\mu}) = s/\sqrt{n}$.

5.2 Point and Interval Estimates

Suppose we got our sample of n basketball players and we used the sample mean, $\hat{\mu}$, as an estimator for μ. Suppose that we found an estimate for μ, e.g. $\hat{\mu} = \bar{Y} = 2.05$m and an estimate for the $\widehat{\text{SE}(\hat{\mu})}$=0.05m (i.e. $\widehat{\text{SE}(\hat{\mu})} = s/\sqrt{n} = 0.05$).

Based on our sample statistics, what can we infer about μ, that is, about the average height of all basketball players?

- Point Estimate: "The (true) average height of all basketball players is $\mu = \bar{Y} = 2.05$m"
- Interval Estimate: Pinpoint how confident we are that the true mean of the height of all basketball players lies within a range around $\bar{Y} = 2.05$m. At the 95% confidence interval and assuming that $\hat{\mu} = \bar{Y} = 2.05$m and $\widehat{\text{SE}(\hat{\mu})}$=0.05m, we have

$$95\% = P(-1.96 < Z < 1.96) = P(-1.96 < \frac{\mu - 2.05}{0.05} < 1.96)$$
$$= P(-1.96 * 0.05 < \mu - 2.05 < 1.96 * 0.05)$$
$$= p(2.05 - 1.96 * 0.05 < \mu < 2.05 + 1.96 * 0.05) = P(1.952 < \mu < 2.148)$$

"With 95% confidence, the true average height of all basketball players will lie between 1.952 and 2.148."

5.3 Hypothesis Testing

Suppose we want to do a hypothesis testing on whether the true average height of all basketball players is $\mu_0 = 1.99$.

1. State the Null Hypothesis $\text{H}_0 : \mu = \mu_0 = 1.99$
2. State the Alternative Hypothesis $\text{H}_1 : \mu \neq \mu_0 \neq 1.99$
3. Construct the Test Statistic under the null (i.e. if H_0 were true)

$$\tau = \frac{\hat{\mu} - \mu_0}{SE(\hat{\mu})} = \frac{2.05 - 1.99}{0.05} = 1.2$$

4. Find how the Test Statistic is distributed under H_0

   We assume that despite σ^2 being unknown we have a large sample ($n > 30$), so that the test statistic (τ) is standard normally distributed

$$\tau \sim SN(0, 1)$$

5. Find the critical values in the tails of the distribution of the test statistic that give the 95% of the distribution: $z^* = \pm 1.96$
6. Apply the decision rule: $\tau = 1.2 < 1.96$. So, we don't reject the null hypothesis and conclude that "at the 95% confidence level, the true mean height of all basketball players is not statistically different from $\mu_0 = 1.99$".

NOTE 1 (Significance level):

P(Type I error)=P(Reject $H_0|H_0$ is true)=5%. If H_0 were true, then any $\hat{\mu} \sim (\mu_0, \sigma^2/n)$ would be consistent with having μ_0 as mean. If the null hypothesis is true, then $\hat{\mu} \sim (\mu_0, \sigma^2/n)$

However, whenever we observe a $|\hat{\mu}|$ that does not lie within 1.96 SE from the true hypothesised mean (or alternatively, $|\tau| > 1.96$) we reject the null, though it is true.

NOTE 2 (Confidence Interval Approach to Hypothesis Testing):

The decision rule is saying that whenever $\hat{\mu}$ lies within 1.96 SE from μ_0 (i.e. whenever $\hat{\mu}$ lies within 1.892 and 2.088) we do not reject the null. This is equivalent to μ_0 being included in the 95% confidence interval of $\hat{\mu}$. The 95% confidence interval was calculated earlier and found to be (1.952, 2.148). The hypothesised value ($\mu_0 = 1.99$) is included in the interval, that that is why we have not rejected the null hypothesis.