Inferential Statistics – Introduction to Hypothesis Testing

Prof. Josefina V. Almeda
College Secretary
School of Statistics
University of the Philippines Diliman
2012

OUTLINE

- Basic Concepts in Hypothesis Testing
- Null and Alternative Hypotheses
- Two Types of Errors
- Types of Hypotheses
- Steps in Conducting a test of Hypothesis

Learning Objectives

After the lesson on Introduction to Hypothesis Testing, the participants should be able to

- State the null and alternative hypotheses;
- Differentiate the two types of errors;
- Learn the types of hypotheses; and
- List the steps in performing a test of hypothesis.

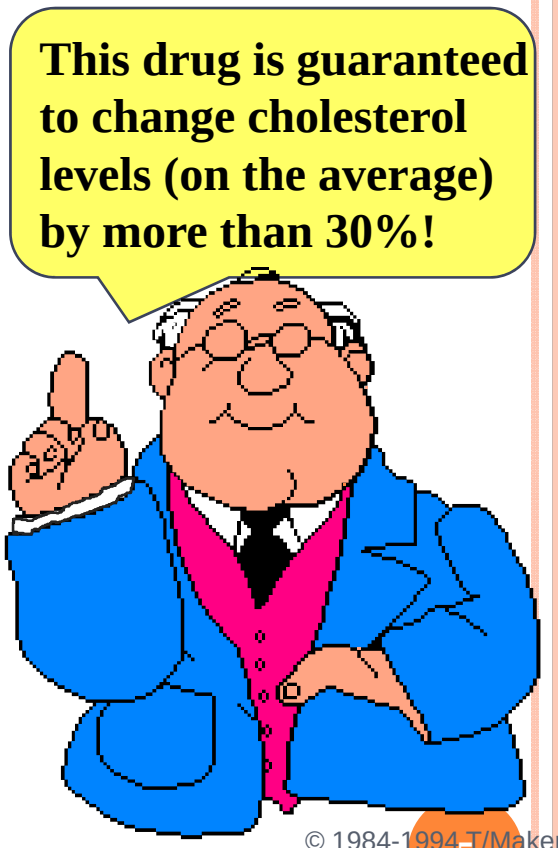
What is Hypothesis Testing?

- Hypothesis testing evaluates a conjecture about some characteristic of the parent population based upon the information contained in the random sample.

- Usually the conjecture concerns one of the unknown parameters of the population.

What is a Hypothesis?

- A hypothesis is a claim or statement about the population parameter
  - Examples of parameters are population mean and population proportion
  - The parameter must be identified before analysis



Example of Hypothesis

- The mean body temperature for patients admitted to elective surgery is not equal to 37.0°C.

Note: The parameter of interest here is μ which is the mean body temperature for patients admitted to elective surgery.

Example of a Hypothesis

- The proportion of registered voters in Quezon City favoring Candidate A exceeds 0.60.

Note: The parameter of interest here is p which is the proportion of registered voters in Quezon City favoring Candidate A.

Components of a Formal Hypothesis Test

Null Hypothesis

- denoted by Ho
- the statement being tested
- it represents what the experimenter doubts to be true
- must contain the condition of equality and must be written with the symbol =

For the mean, the null hypothesis will be

- Ho: μ = some value

- Note: the hypothesized value of μ can be obtained from previous studies or from knowledge of the population

Example of Null Hypothesis

- The null hypothesis corresponding to the common belief that the mean body temperature is 37°C is expressed as

Ho: $\mu = 37\ ^{o}C$

- We test the null hypothesis directly in the sense that we assume it is true and reach a conclusion to either reject Ho or fail to reject Ho.

Alternative Hypothesis

- denoted by Ha
- Is the statement that must be true if the null hypothesis is false
- the operational statement of the theory that the experimenter believes to be true and wishes to prove
- Is sometimes referred to as the research hypothesis

For the mean, the alternative hypothesis will be stated in only one of three possible forms:

- Ha: $\mu \neq$ some value
- Ha: $\mu >$ some value
- Ha: $\mu <$ some value

Note: Ha is the opposite of Ho. For example, if Ho is given as $\mu = 37.0$, then it follows that the alternative hypothesis is given by Ha: $\mu \neq 37.0$.

Example in Stating your Hypothesis

If you believe that your brand of refrigerator lasts longer than the mean of 14 years for other brands, state the claim that $\mu > 14$, where μ is the mean life of your refrigerators.

Ho: $\mu = 14$ vs. Ha: $\mu > 14$

Some Notes:

- In this context of trying to support the goal of the research, the alternative hypothesis is sometimes referred to as the *research hypothesis*.
- Also in this context, the null hypothesis is assumed true for the purpose of conducting the hypothesis test, but it is hoped that the conclusion will be rejection of the null hypothesis so that the research hypothesis is supported.

What is a Test of Significance?

- A test of significance is a problem of deciding between the null and the alternative hypotheses on the basis of the information contained in a random sample.

- The goal will be to reject Ho in favor of Ha, because the alternative is the hypothesis that the researcher believes to be true. If we are successful in rejecting Ho, we then declare the results to be “significant”.

Two Types of Errors

- Type I Error
- Type II Error

Type I Error

- The mistake (error) of rejecting the null hypothesis when it is true.
- It is not a miscalculation or procedural misstep; it is an actual error that can occur when a rare event happens by chance.
- The probability of rejecting the null hypothesis when it is true is called the significance level (α).
- The value of α is typically predetermined, and very common choices are $\alpha = 0.05$ and $\alpha = 0.01$.

Examples of Type I Error

1. The mistake of rejecting the null hypothesis that the mean body temperature is 37.0 when that mean is really 37.0.

2. BFaD allows the release of an ineffective medicine

Type II Error

- The mistake of failing to reject the null hypothesis when it is false.
- The symbol β (beta) is used to represent the probability of a type II error.

Examples of Type II Errors

1. The mistake of failing to reject the null hypothesis ($\mu = 37.0$) when it is actually false (that is, the mean is not 37.0).

2. BFaD does not allow the release of an effective drug.

Summary of Possible Decisions in Hypothesis Testing

		True Situation: The null hypothesis is true.	True Situation: The null hypothesis is false.
Decision	We decide to reject the null hypothesis.	**TYPE I error** *(rejecting a true null hypothesis)*	CORRECT decision
	We fail to reject the null hypothesis.	CORRECT decision	**TYPE II error** *(failing to reject a false null hypothesis)*

Analogy to Decisions in Hypothesis Testing

Trial

	The Truth	
Verdict	Innocent	Guilty
Innocent	**Correct**	**Error**
Guilty	**Error**	**Correct**

Controlling Type I and Type II Errors

- The experimenter is free to determine α. If the test leads to the rejection of Ho, the researcher can then conclude that there is sufficient evidence supporting Ha at α level of significance.

- Usually, β is unknown because it's hard to calculate it. The common solution to this difficulty is to "withhold judgment" if the test leads to the failure to reject Ho.

- α and β are inversely related. For a fixed sample size n, as α decreases β increases.

Controlling Type I and Type II Errors

- In almost all statistical tests, both α and β can be reduced by increasing the sample size.

- Because of the inverse relationship of α and β, setting a very small α should also be avoided if the researcher cannot afford a very large risk of committing a Type II error.

Controlling Type I and Type II Errors

- The choice of α usually depends on the consequences associated with making a Type I error.

Common Choices of α	Consequences of Type I error
0.01 or smaller	very serious
0.05	moderately serious
0.10	not too serious

Controlling Type I and Type II Errors

- The usual practice in research and industry is to determine in advance the values of α and n, so the value of β is determined.
- Depending on the seriousness of a type I error, try to use the largest α that you can tolerate.
- For type I errors with more serious consequences, select smaller values of α . Then choose a sample size n as large as is reasonable, based on considerations of time, cost, and other such relevant factors.

Example to illustrate Type I and Type II Errors

Consider M&Ms (produced by Mars, Inc.) and Bufferin brand aspirin tablets (produced by Bristol-Myers Products).

The M&M package contains 1498 candies. The mean weight of the individual candies should be at least 0.9085 g., because the M&M package is labeled as containing 1361 g.

Example to illustrate Type I and Type II Errors

The Bufferin package is labeled as holding 30 tablets, each of which contains 325 mg of aspirin.

Because M&Ms are candies used for enjoyment whereas Bufferin tablets are drugs used for treatment of health problems, we are dealing with two very different levels of seriousness.

Example to illustrate Type I and Type II Errors

If the M&Ms don't have a population mean weight of 0.9085 g, the consequences are not very serious, but if the Bufferin tablets don't have a mean of 325 mg of aspirin, the consequences could be very serious.

Example to illustrate Type I and Type II Errors

If the M&Ms have a mean that is too large, Mars will lose some money but consumers will not complain.

In contrast, if the Bufferin tablets have too much aspirin, Bristol-Myers could be faced with consumer lawsuits.

Example to illustrate Type I and Type II Errors

Consequently, in testing the claim that $\mu = 0.9085$ g for M&Ms, we might choose $\alpha = 0.05$ and a sample size of $n = 100$.

In testing the claim of $\mu = 325$ mg for Bufferin tablets, we might choose $\alpha = 0.01$ and a sample size of $n = 500$.

The smaller significance level α and large sample size n are chosen because of the more serious consequences associated with the commercial drug.

Types of Tests

- **Two-tailed Test.** If we are primarily concerned with deciding whether the true value of a population parameter is different from a specified value, then the test should be two-tailed. For the case of the mean, we say Ha: $\mu \neq \mu_0$.
- **Left-tailed Test.** If we are primarily concerned with deciding whether the true value of a parameter is less than a specified value, then the test should be left-tailed. For the case of the proportion, we say Ha: $P < P_0$.
- **Right-tailed Test.** If we are primarily concerned with deciding whether the true value of a parameter is greater than a specified value, then we should use the right-tailed test. For the case of the standard deviation, we say Ha: $\sigma > \sigma_0$.

The p-value - the smallest level of significance at which Ho will be rejected based on the information contained in the sample

Form of Decision Rule

(based on the p-value)

Reject Ho if the p-value is less than or equal to the level of significance (α).

Example of Making Decisions Using the p-value

If the level of significance $\alpha=0.05$,

p-value	Decision
0.01	Reject Ho.
0.05	Reject Ho.
0.10	Do not reject Ho

Conclusions in Hypothesis Testing

1. Fail to reject the null hypothesis Ho.
2. Reject the null hypothesis Ho.

Notes:

- Some texts say "accept the null hypothesis" instead of "fail to reject the null hypothesis."
- Whether we use the term accept or fail to reject, we should recognize that we are not proving the null hypothesis; we are merely saying that the sample evidence is not strong enough to warrant rejection of the null hypothesis.

Determine the testing procedure and methods of analysis (responsibility of the statistician).

3. Decide on the type of data to be collected and choose an appropriate test statistic and testing procedure.

Determine the testing procedure and methods of analysis (responsibility of the statistician).

3. Decide on the type of data to be collected and choose an appropriate test statistic and testing procedure.

CON'T OF STEPS IN HYPOTHESIS TESTING

4. **State the decision rule.**

5. Obtain the p-value.

6. Make the conclusion.

   If the p-value is less than or equal to α, reject the Ho at α level of significance and support the alternative hypothesis.

   If the p-value is greater than α, do not reject the Ho. There is no sufficient sample evidence to reject the null hypothesis.