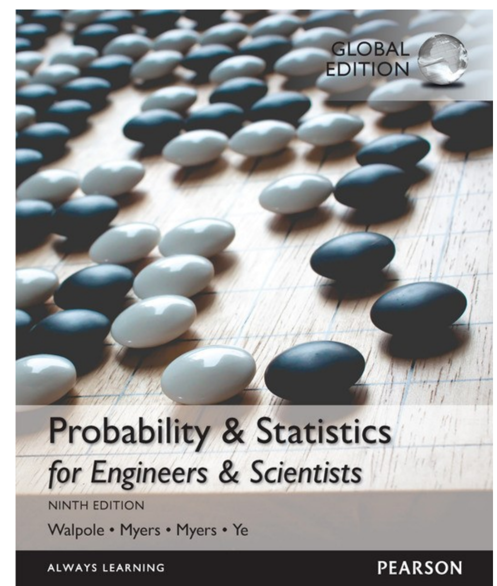
Chapter 10

One- and Two-Sample Tests of Hypotheses

Statistical Hypotheses: General Concepts

School of Computing, Gachon Univ.
Joon Yoo



가천대학교 AI·소프트웨어학부
Gachon University School of Computing

10.1 General Concepts

Statistical Inference (recall)

- Estimation (Last chapter)
  - Taking a random sample from the distribution to elicit some information about the unknown parameter θ.
  - Example
    - A candidate for public office may wish to estimate the true proportion of voters favoring him by obtaining the opinions from a random sample of 100 eligible voters

- Hypothesis Testing (This chapter)
  - We do not attempt to estimate a parameter, but instead we try to arrive at a correct decision about a pre-stated hypothesis.
  - Example
    - One is interested in finding out whether brand A floor wax is more scuff-resistant than brand B floor wax. He or she might hypothesize that brand A is better than brand B and, after proper testing, accept or reject this hypothesis.

- Instead of making an estimate about a population parameter, you'll learn how to test a conjecture about a parameter.

conjecture (추측)

- Example 1
  - Suppose that you work for Gallup and are asked to test a claim that the proportion of eligible American voters who support Barack Obama is p = 0.47.
  - To test the claim, you take a random sample of n = 1200 eligible voters and find 594 of them support Barack Obama. Your sample statistic is $\hat{p} = 0.495$.
  - Is the sample statistic identical enough to your claim (p = 0.47) to decide that the claim is true, or different enough from the claim (p = 0.47) to decide that the claim is false?

- Example 2
  - A medical researcher may decide on the basis of experimental evidence whether coffee drinking increases the risk of cancer in humans.
  - A sociologist might wish to collect appropriate data to enable him or her to decide whether a person's blood type and eye color are independent variables.

Statistical hypotheses

- In each case, the conjecture can be put in the form of a statistical hypothesis*

- Definition 10.1:
  - **A statistical hypothesis is an assertion or conjecture concerning (parameters of) one or more populations**

- *Truth* is never known unless we examine the entire population.

*가설

Hypothesis Test

- Hypothesis testing*: **accept** or **reject** a hypothesis based on the sample information

- As always, we take a random *sample* from the population.
- Evidence from the sample that is inconsistent with the stated hypothesis leads to a rejection of the hypothesis.
  - implies that “the sample evidence” refutes it.

*가설검정

The Role of Probability in Hypothesis Testing

Example:

- Hypothesis:

  A fraction *p= 0.1* of a production is defective

- Sample of 100 units, 12 defective

refute (논박하다)

Can we refute the hypothesis based on this result ?

Can we refute that p = 0.12 or p = 0.15 based on this result ?

The Role of Probability in Hypothesis Testing

Example:

- Hypothesis:

  A fraction $p = 0.1$ of a production is defective

- If there are **20 defective units** in a sample of 100 units.
  → Hypothesis that $p=0.1$ is unlikely
- Why?
  - If $p=0.1$, **probability** of observing 20 defective units is **0.002**
- The rejection of a hypothesis means that
  - there is a small probability of obtaining the sample information observed
  - the evidence from the sample refutes* it.
- With the resulting small risk of wrong conclusion, it would seem safe to reject the hypothesis that p=0.10.

**refute* (논박하다)

Rejection of a Hypothesis

- Rejection rules out* the hypothesis
- Acceptance, or rather failing to reject, does not rule out other possibilities.

The <u>firm conclusion</u> is established when a hypothesis is rejected.

Sure

- The formal statement of a hypothesis is often influenced by the structure of the probability of a wrong conclusion.
- If the scientist is interested in strongly supporting a contention, he or she hopes to arrive at the contention in the form of rejection a hypothesis.
- Rejecting a hypothesis is stronger than failing to reject it.

**rule out* (배제하다)

Example 1

- If the medical researcher wishes to show strong evidence in favor of the contention that coffee drinking increases the risk of cancer,
- the hypothesis tested should be of the form ”there is no increase in cancer risk produced by drinking coffee.” As a result, the contention is reached via a rejection.

Example 2

- Similarly, to support the claim that one kind of gauge is more accurate than another,
- the engineer tests the hypothesis that there is no difference in the accuracy of the two kinds of gauges.

- The foregoing implies that when the data analyst formalizes experimental evidence on the basis of hypothesis testing, the **formal statement** of hypothesis is very important.

The Null and Alternative Hypotheses

- **Null hypothesis H_0:**
  - The hypothesis we wish to test, which is denoted by H_0
- **Alternative hypothesis H_1:**
  - **The rejection of H_0 leads to acceptance of this hypothesis**, denoted by H_1
- H_1 is the question to be answered, or the theory to be tested
- H_0 nullifies or opposes H_1, and often is the logical complement of H_0

Arriving conclusions

- ***reject*** $\boldsymbol{H_0}$: in favor of H_1 because of sufficient evidence in the data
- ***fail to reject*** $\boldsymbol{H_0}$: because of insufficient evidence in the data.
  - We fail to *accept* H_1

arrive at one of two conclus-ions

- Note: conclusions do not involve a formal and literal "***accept*** $\boldsymbol{H_0}$"

Arriving conclusions

- In our binomial example, the practical issue may be a concern that the historical defective probability of 0.10 no longer is true.
- Indeed, the conjecture may be that p exceeds 0.10. We may then state:

$$H_0 : p = 0.1$$
$$H_1 : p > 0.1$$

Now 12 defective items out of 100 does not refute $p = 0.10$, so the conclusion is "fail to reject H_0." However, if the data produce 20 out of 100 defective items, then the conclusion is "reject H_0" in favor of H_1: $p > 0.10$.

Fails to Reject: Example

Though the applications of hypothesis testing are quite abundant in scientific and engineering work, perhaps the best illustration for a novice lies in the predicament encountered in a jury trial. The null and alternative hypotheses are

H_0: defendant is innocent,

H_1: defendant is guilty.

The indictment comes because of suspicion of guilt. The hypothesis H_0 (the status quo) stands in opposition to H_1 and is maintained unless H_1 is supported by evidence "beyond a reasonable doubt." However, "failure to reject H_0" in this case does not imply innocence, but merely that the evidence was insufficient to convict. So the jury does not necessarily *accept* H_0 but *fails to reject* H_0.

혐의없음
증거불충분

10.2 Testing a Statistical Hypothesis

Testing a Statistical Hypothesis

- A certain type of cold vaccine is known to be only 25% effective after a period of 2 years.
- We want to determine if a new kind of vaccine is effective for a longer period of time.

- Experiment:
  - Choose 20 people at random and inoculate them with the new vaccine (In actual situations we need thousands of people)
  - If more than **8** remain healthy after 2 years, we conclude that the new vaccine is better.

Example (cont.)

- The number 8 seems arbitrary, but reasonable
  - represents a modest gain over the 5 people (25%) who could be expected to receive protection if the 20 people had been inoculated with old vaccine.
- We are testing the *null hypothesis* that
  - The new vaccine is <u>equally effective</u> after 2 years as the former one.
- The *alternative hypothesis* is that
  - The new vaccine is in fact superior.

Example (cont.)

- This is equal to testing the hypothesis that the binomial parameter for the probability of a success on a given trial is $p = 0.25$, against the alternative that $p > 0.25$.

The new vaccine is equally effective after 2 years as the former one.

$$H_0 : p = 0.25$$

$$H_1 : p > 0.25$$

The new vaccine is in fact superior.

The Test Statistic

- The **test statistic** is the observed statistic on which we base our decision.
  - In this case, it is **X** (0~20), the number of healthy people after two years
  - divided into two groups: those numbers less than or equal to 8 ($X \le 8$) and those greater than 8 ($X > 8$).
- The values of X that makes us reject the null hypothesis constitute the **critical region** ($X>8$)
- The last number we observe *before* passing into the critical region is the **critical value**
  - In this case, it is 8.

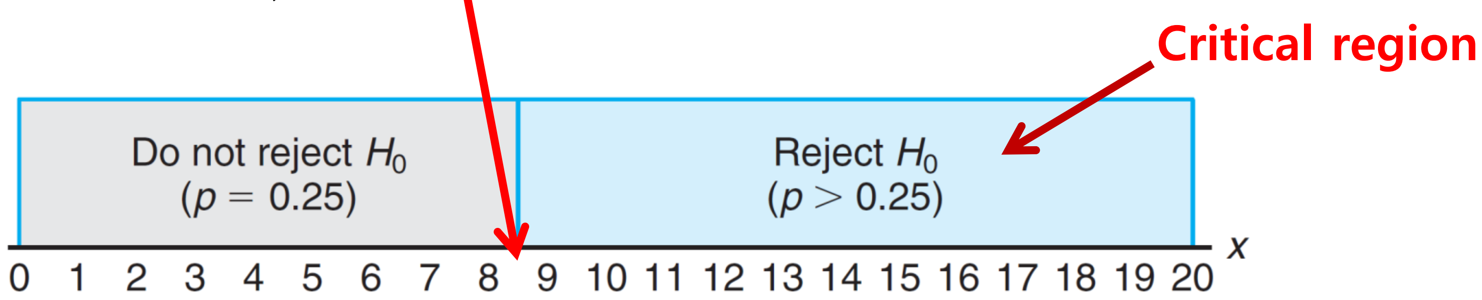


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

Types of Error

- This decision procedure could lead to either of two **wrong conclusions**

H_0 True

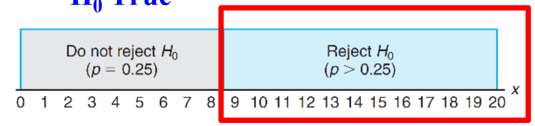


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

> For instance, the new vaccine may be no better than the one now in use (H_0 true) and yet, in this particular randomly selected group of individuals, more than 8 surpass the 2-year period without contracting the virus.

Definition 10.2: Rejection of the null hypothesis when it is true is called a **type I error**.

- **Type I error:**
  - Rejecting H_0 in favor of H_1 when, in fact, H_0 is true.
  - The probability of a type I error, also called **level of significance**, is denoted by **α**.

Types of Error

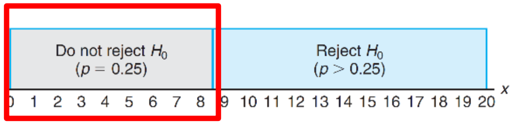


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

A second kind of error is committed if 8 or fewer of the group surpass the 2-year period successfully and we are unable to conclude that the vaccine is better when it actually is better (H_1 true). Thus, in this case, we fail to reject H_0 when in fact H_0 is false. This is called a **type II error**.

Definition 10.3: Nonrejection of the null hypothesis when it is false is called a **type II error**.

- **Type II error:**
  - Failing to reject H_0 when, in fact, H_0 is false.
  - The probability of type II error, denoted by **β**, is impossible to compute unless we have a specific H_1

Possible Situations

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

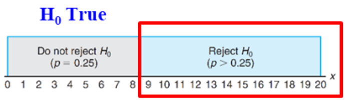


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

Probability of committing Type I error (=level of significance α)

$$\alpha = P(\text{type I error}) = P\left(X > 8 \text{ when } p = \frac{1}{4}\right) = \sum_{x=9}^{20} b\left(x; 20, \frac{1}{4}\right)$$

$$= 1 - \sum_{x=0}^{8} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.9591 = 0.0409.$$

We say, "the null hypothesis, p= ¼ (or 0.25), is being tested at the α =0.0409 level of significance."

α =0.0409 is very small, therefore, a Type 1 error is unlikely. Consequently, it would be unusual for more than 8 individuals to remain immune to a virus for a 2-year period using a new vaccine (p=0.25) equivalent to the one now on the market.

Computing Type II Error

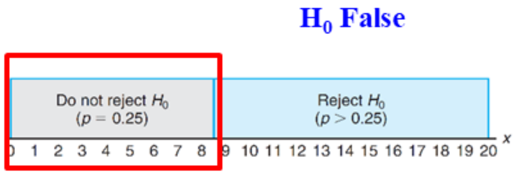


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

- Assumption
  - H_0: Null hypothesis is p = 0.25 (1/4)
  - H_1: As alternative hypothesis, use a specific value for p, such as p = 0.5 (1/2)
- Then, we get

$$\beta = P(\text{type II error}) = P\left(X \leq 8 \text{ when } p = \frac{1}{2}\right)$$

$$= \sum_{x=0}^{8} b\left(x; 20, \frac{1}{2}\right) = 0.2517.$$

  - This is a rather high probability (0.2517), indicating a test procedure in which it is quite likely that we shall reject the new vaccine when, in fact, it is superior to what is now in use.

Computing Type II Error

- It is possible that the director of the testing program is willing to make a type II error if the more expensive vaccine is not significantly superior. In fact, the only time he wishes to guard against the type II error is when the true value of p is at least 0.7.
- Assumption
  - H_0: Null hypothesis is p = 0.25 (1/4)
  - H_1: As alternative hypothesis, use a specific value for p, such as p = 0.7 (7/10)
- Then, we get
  - if p=0.7,

$$\beta = P(\text{type II error}) = P(X \le 8 \text{ when } p = 0.7)$$

$$= \sum_{x=0}^{8} b(x; 20, 0.7) = 0.0051.$$

  - it is extremely unlikely that the new vaccine would be rejected when it was 70% effective after a period of 2 years.

The Role of α, β , and Sample Size

- Ideally, both types of errors (α, β) should be small
- For some applications, one type of error might be more important than the other
- How to change α and β ?
  - Either change the critical value
    - Usually decreases one type of error while increasing the other
  - Or change the sample size
    - Increasing the sample size reduces *both* types of error

The role of α, β and the Sample Size

- In our example
  - Change **critical value** from 8 to 7
    - α increases: 0.0409 → 0.1018
    - β decreases: 0.2517 → 0.1316

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

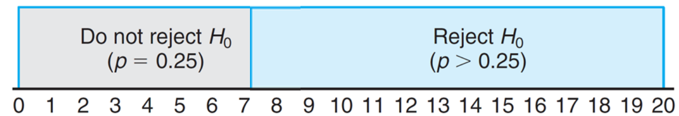


Figure 10.1: Decision criterion for testing $p = 0.25$ versus $p > 0.25$.

$$\alpha = \sum_{x=8}^{20} b\left(x; 20, \frac{1}{4}\right) = 1 - \sum_{x=0}^{7} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.8982 = 0.1018$$

and

$$\beta = \sum_{x=0}^{7} b\left(x; 20, \frac{1}{2}\right) = 0.1316.$$

  - By adopting a new decision procedure, we have reduced the probability of committing a type II error β at the expense of increasing the probability of committing a type I error α.

The role of α, β and the Sample Size

- In our example
  - Change **sample size** from 20 to 100
  - The critical value is now 36.
    - α decreases: $0.0409 \rightarrow 0.0039$
    - β decreases: $0.2517 \rightarrow 0.0035$
  - Refer to pages 344-345 for details of above
- These concepts can be equally well applied to continuous random variables.

Illustration with a Continuous Random Variable

The concepts discussed here for a discrete population can be applied equally well to continuous random variables. Consider the null hypothesis that the average weight of male students in a certain college is 68 kilograms against the alternative hypothesis that it is unequal to 68. That is, we wish to test

$$H_0\colon\ \mu = 68,$$
$$H_1\colon\ \mu \neq 68.$$

The alternative hypothesis allows for the possibility that $\mu < 68$ or $\mu > 68$.

A sample mean that falls close to the hypothesized value of 68 would be considered evidence in favor of H_0. On the other hand, a sample mean that is considerably less than or more than 68 would be evidence inconsistent with H_0 and therefore favoring H_1. The sample mean is the test statistic in this case. A critical region for the test statistic might arbitrarily be chosen to be the two intervals $\bar{x} < 67$ and $\bar{x} > 69$. The nonrejection region will then be the interval $67 \leq \bar{x} \leq 69$. This decision criterion is illustrated in Figure 10.4.

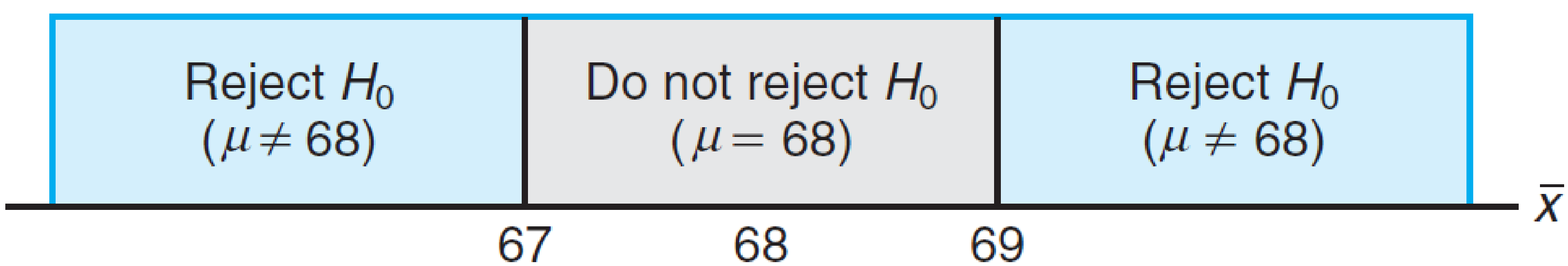


Figure 10.4: Critical region (in blue).

Assume the standard deviation of the population of weights to be $\sigma = 3.6$. For large samples, we may substitute s for σ if no other estimate of σ is available. Our decision statistic, based on a random sample of size $n = 36$, will be $\bar{X}$, the most efficient estimator of μ. From the Central Limit Theorem, we know that the sampling distribution of $\bar{X}$ is approximately normal with standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3.6/6 = 0.6$.

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

H_0: $\mu = 68$,

H_1: $\mu \neq 68$.

Type I error: Rejecting H_0 in favor of H_1 when, in fact, H_0 is true.

$$\alpha = P(\bar{X} < 67 \text{ when } \mu = 68) + P(\bar{X} > 69 \text{ when } \mu = 68).$$

The z-values corresponding to $\bar{x}_1 = 67$ and $\bar{x}_2 = 69$ when H_0 is true are

$$z_1 = \frac{67 - 68}{0.6} = -1.67 \quad \text{and} \quad z_2 = \frac{69 - 68}{0.6} = 1.67.$$

Therefore,

$$\alpha = P(Z < -1.67) + P(Z > 1.67) = 2P(Z < -1.67) = 0.0950.$$

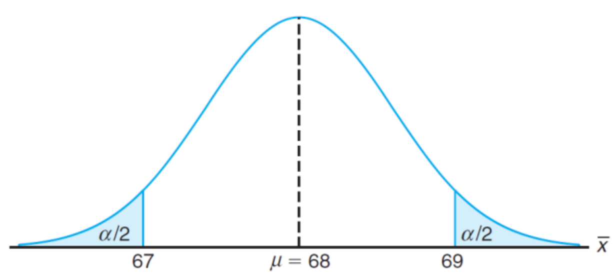


Figure 10.5: Critical region for testing $\mu = 68$ versus $\mu \neq 68$.

Type II error: Failing to reject H_0 when, in fact, H_0 is false.

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

H_0: $\mu = 68$,

H_1: $\mu \neq 68$.

The reduction in α is not sufficient by itself to guarantee a good testing procedure. We must also evaluate β for various alternative hypotheses. If it is important to reject H_0 when the true mean is some value $\mu \geq 70$ or $\mu \leq 66$, then the probability of committing a type II error should be computed and examined for the alternatives $\mu = 66$ and $\mu = 70$. Because of symmetry, it is only necessary to consider the probability of not rejecting the null hypothesis that $\mu = 68$ when the alternative $\mu = 70$ is true. A type II error will result when the sample mean $\bar{x}$ falls between 67 and 69 when H_1 is true. Therefore, referring to Figure 10.6, we find that

$$\beta = P(67 \leq \bar{X} \leq 69 \text{ when } \mu = 70).$$

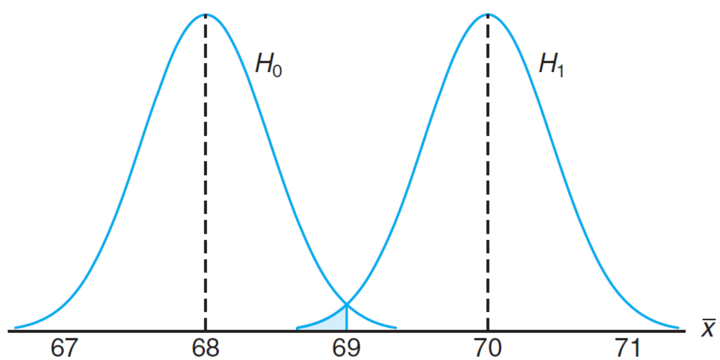


Figure 10.6: Probability of type II error for testing $\mu = 68$ versus $\mu = 70$.

The z-values corresponding to $\bar{x}_1 = 67$ and $\bar{x}_2 = 69$ when H_1 is true are

$$z_1 = \frac{67 - 70}{0.45} = -6.67 \quad \text{and} \quad z_2 = \frac{69 - 70}{0.45} = -2.22.$$

Therefore,

$$\begin{aligned}\beta = P(-6.67 < Z < -2.22) &= P(Z < -2.22) - P(Z < -6.67)\\ &= 0.0132 - 0.0000 = 0.0132.\end{aligned}$$

If the true value of μ is the alternative $\mu = 66$, the value of β will again be 0.0132. For all possible values of $\mu < 66$ or $\mu > 70$, the value of β will be even smaller when $n = 64$, and consequently there would be little chance of not rejecting H_0 when it is false.

The probability of committing a type II error increases rapidly when the true value of μ approaches, but is not equal to, the hypothesized value. Of course, this is usually the situation where we do not mind making a type II error. For example, if the alternative hypothesis $\mu = 68.5$ is true, we do not mind committing a type II error by concluding that the true answer is $\mu = 68$. The probability of making such an error will be high when $n = 64$. Referring to Figure 10.7, we have

$$\beta = P(67 \leq \bar{X} \leq 69 \text{ when } \mu = 68.5).$$

The z-values corresponding to $\bar{x}_1 = 67$ and $\bar{x}_2 = 69$ when $\mu = 68.5$ are

$$z_1 = \frac{67 - 68.5}{0.45} = -3.33 \quad \text{and} \quad z_2 = \frac{69 - 68.5}{0.45} = 1.11.$$

Therefore,

$$\begin{aligned}\beta = P(-3.33 < Z < 1.11) &= P(Z < 1.11) - P(Z < -3.33) \\ &= 0.8665 - 0.0004 = 0.8661.\end{aligned}$$

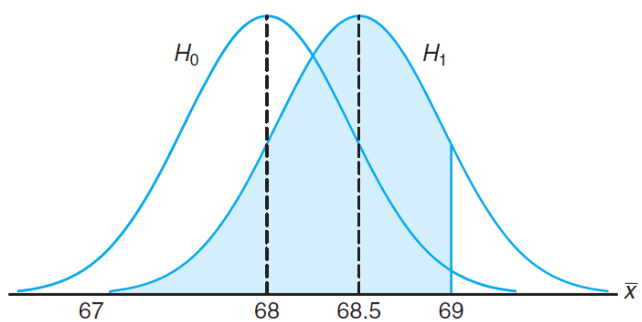


Figure 10.7: Type II error for testing $\mu = 68$ versus $\mu = 68.5$.

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

H_0: $\mu = 68$,

H_1: $\mu \neq 68$.

Important Properties of a Test of Hypothesis

- 1- The type I error and type II error are related.
  - A decrease in the probability of one generally results in an increase in the probability of the other

- 2- The size of the critical region, and therefore the probability of committing a type I error, can always be reduced by adjusting the critical values

Important Properties of a Test of Hypothesis

- 3- An increase in the sample size will reduce both types of error simultaneously

- 4- If the null hypothesis is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The greater the distance between the true value and the hypothesized value, the smaller β will be.

The Power of a Test

The **power** of a test : 검정력

- Definition:
  - The **power** of a test is the probability of rejecting H_0 given that a specific alternative is true
    - Which is $1-\beta = P(\text{Reject } H_0 | H_1 \text{ is True}).$

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

- Different kinds of tests are compared by contrasting power properties.
  - To increase the power of a test, either increase α (=decrease β), or increase sample size

Example

- In a case of testing $H_0 : \mu = 68$ and $H_1 : \mu \neq 68$.

- if μ is truly 68.5, probability of a type II error is given by $\beta = 0.8661$.
- The power of the test is $1 - 0.8661 = 0.1339$.
- the test as described will *properly reject H_0 only 13.39% of the time*.
  - Not a good test!

One- and Two-Tailed Tests

- One-tailed test
  - A test of any hypothesis where the alternative is **one sided**, such as
    - $H_0 : \theta = \theta_0$
    - $H_1 : \theta > \theta_0$ or $H_1 : \theta < \theta_0$ where critical region is not split

$$H_0:\ p = 0.25,$$
$$H_1:\ p > 0.25.$$

- Two-tailed test
  - A test of any hypothesis where the alternative is **two sided**, such as
    - $H_0 : \theta = \theta_0$
    - $H_1 : \theta \neq \theta_0$ where critical region is split into two parts
      - » H_1 *states* $\theta \neq \theta_0$ states that either $\theta < \theta_0$ or $\theta > \theta_0$.

$$H_0:\ \mu = 68,$$
$$H_1:\ \mu \neq 68.$$

One- and Two-Tailed Tests

- Whether one sets up a one-tailed ? or a two-tailed test ?

- Whether one sets up a one-tailed ? or a two-tailed test will depend on the conclusion to be drawn if H_0 is rejected.

Choosing Null and Alternative Hypotheses

- Example 10.1
  - A manufacturer of a certain brand of rice cereal claims that the average saturated fat content does not exceed 1.5 mg.
    - State the null and alternative hypotheses to be used in testing this claim and determine where the critical region is located.

- Solution
  - The claim should be rejected only if the average is greater than 1.5 mg and should not be rejected if average is less than or equal to 1.5 mg.
    - $H_0 : \mu = 1.5mg$
    - $H_1 : \mu > 1.5mg$ - one-tailed
  - Note that the nonrejection of H_0 does not rule out values less than 1.5mg. The critical region lies entirely in the right tail of the distribution.

- Example 10.2
  - A real estate agent claims that 60% of all private residences being built today are 3-bedroom homes. To test this claim, a large sample of new residences is inspected; the proportion of these homes with 3 bedrooms is recorded and used as our test statistic.
    - State the null and alternative hypotheses to be used in testing this claim and determine where the critical region is located.

- Solution
  - We reject if the test statistic is significantly higher or lower than p=0.6.
    - H_0 : p = 0.6
    - H_1 : p ≠ 0.6
  - The alternative hypothesis implies a two-tailed test, where the critical region is symetrically split into two.

https://www.psycom.net/bipolar-questions-answers