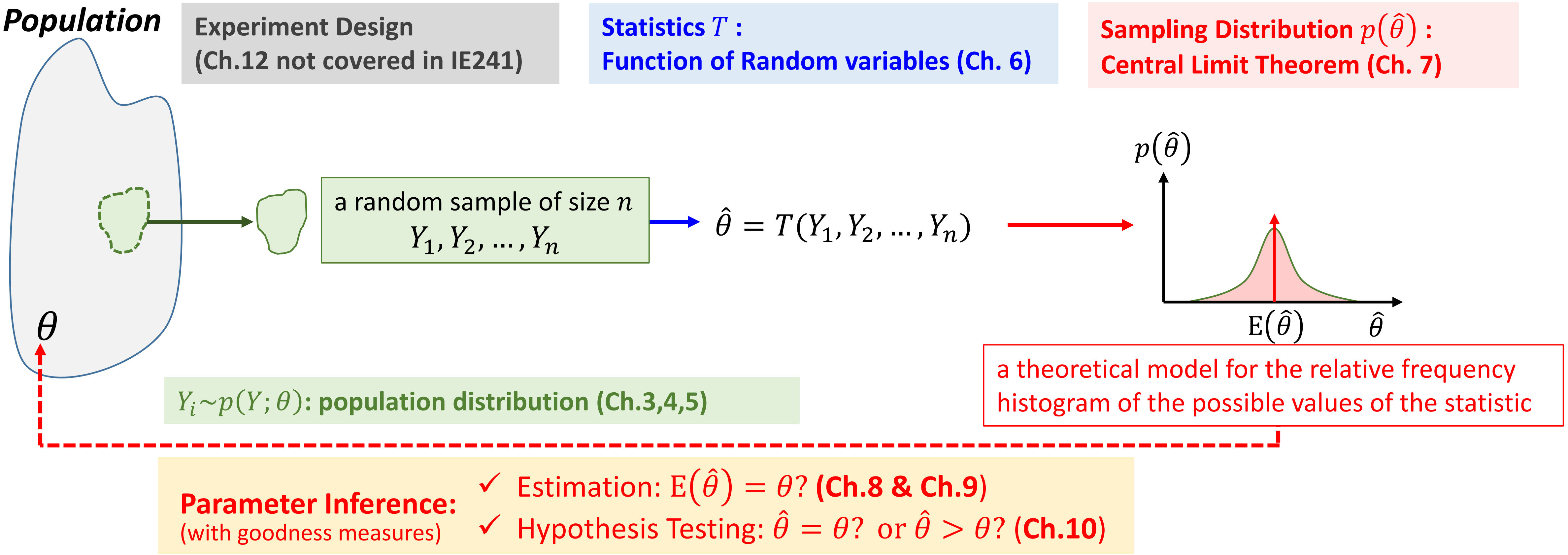
CHAPTER 10
Hypothesis Testing

Motivation

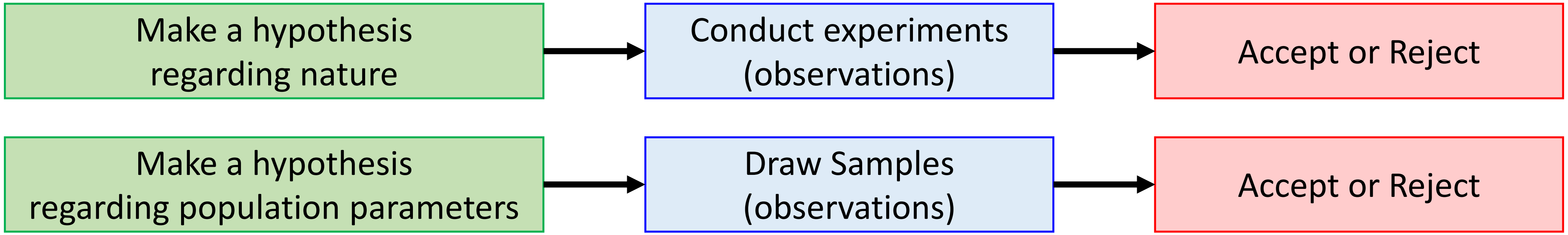


- The objective of statistics often is to make inferences about unknown population parameters based on information contained in sample data.
- Two different ways of inference are:
  - ✓ Estimates of the respective parameters (Chapter 8, 9)
  - ✓ Tests of hypotheses about their values (Chapter 10)

Motivation

The formal procedure for hypothesis testing is similar to the scientific method.

1. The scientist poses a hypothesis concerning one or more population parameters—that they equal specified values
2. She then samples the population and compares her observations with the hypothesis
   - ✓ If the observations ***disagree*** with the hypothesis, the scientist **rejects** it.
   - ✓ If not, the scientist concludes either that the hypothesis is true or that the sample did not detect the difference between the real and hypothesized values of the population parameters.



Motivation

- Hypothesis tests are conducted in all fields in which theory can be tested against observation.
  - A medical researcher may hypothesize that a new drug is more effective than another in combating a disease
  - A quality control engineer may hypothesize that a new assembly method produces only 5% defective items.
  - An educator may claim that two methods of teaching reading are equally effective
  - A political candidate may claim that a plurality of voters favor his election.

Roles of Statistics in Hypothesis Testings

- What is the role of statistics in testing hypotheses?
- Testing a hypothesis requires making a decision when comparing the observed sample with theory.
    - How do we decide whether the sample disagrees with the scientist's hypothesis?
    - When should we reject the hypothesis?
    - when should we accept it?
    - when should we withhold judgment?
    - What is the probability that we will make the wrong decision ?
    - What function of the sample measurements (test statistics) should be used for decision?

The answers to these questions are contained in a study of statistical hypothesis testing.

Elements of a Statistical Test

- Support for one theory is obtained by showing lack of support for its converse – in a sense, a proof by contradiction

Hypothesis

Hypotheses (H_0 and H_a)

- ✓ **Theory** (H_a: alternative hypothesis) : Jones is not favored by more than 50% of the electorate
- ✓ **Converse** (H_0: null hypothesis) : Jones claim will gain more than 50% of vote in a city

Observation

Test statistic

$n = 15$ voters are randomly selected from the city and Y (the number favoring Jones) is recorded as 0

Decision making

Rejection region

P($Y = 0 | H_0$) is too small (highly improbable)

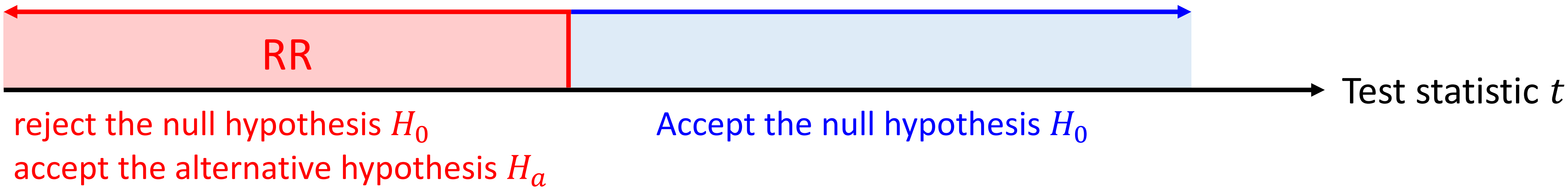
→ Reject H_0 and accept H_a

- **The Elements of a Statistical Test**

What would you like to challenge? → Null hypothesis, H_0
What would you like to support? → Alternative hypothesis, H_a
What statistics of sample measurements are you going to use? → Test statistics T
What criterion are you going to use to make decision? → Rejection region RR

Rejection Region

- The rejection region RR specifies the values of the test statistic for which the null hypothesis is to be rejected in favor of the alternative hypothesis.
  - If for a particular sample, if the computed value of the test statistic falls in the rejection region RR, we reject the null hypothesis H_0 and accept the alternative hypothesis H_a.
  - If the value of the test statistic does not fall in to the RR, we accept H_0.



- For election example small values of Y would lead us to reject H_0. Therefore, one rejection region that we might want to consider is the set of all values of Y less than or equal to 2.
  - ✓ We will use the notation $\mathrm{RR} = \{y : y \le 2\}$—or, more simply, $\mathrm{RR} = \{y \le 2\}$— to denote this rejection region.

Measure a goodness of a test

- Finding a good rejection region for a statistical test is an interesting problem
- We intuitively choose the rejection region as $\mathrm{RR} = \{y \leq k\}$, what k should be used?
- We need to criteria to measure goodness of a specified rejection region
- For any fixed rejection region (determined by a particular value of k), two types of errors can be made in reaching a decision.
  - we can decide in favor of H_a when H_0 is true (make a *type I error*), or
  - we can decide in favor of H_0 when H_a is true (make a *type II error*).

	H_0 is True	H_0 is False
Reject H_0	*type I error* $P(\textit{type I error}) = \alpha$ **(level of test)**	Correct Decision
Accept H_0	Correct Decision	*type II error* $P(\textit{type II error}) = \beta$

Definition

EXAMPLE 10.1

For Jones's political poll, $n = 15$ voters were sampled. We wish to test $H_0\colon\ p = .5$ against the alternative, $H_a\colon\ p < .5$. The test statistic is Y, the number of sampled voters favoring Jones. Calculate α if we select $RR = \{y \le 2\}$ as the rejection region.

Definition

SOLUTION 10.1

By definition,

$$\alpha = P(type\ I\ error) = P(rejecting\ H_0\ when\ H_0\ is\ true)$$
$$= P(value\ of\ test\ statistic\ is\ in\ RR\ when\ H_0\ is\ true)$$
$$= P(Y \leq 2\ when\ p = 0.5)$$

Observe that Y is a binomial random variable with $n = 15$. If H_0 is true, $p = 0.5$ and we obtain

$\alpha = \Sigma_{y=0}^{2} \binom{15}{y} (0.5)^y\ (0.5)^{15-y} = .004$ using Table 1, Appendix 3.

Example

EXAMPLE 10.2

Refer to Example 10.1. Is our test equally good in protecting us from concluding that Jones is a winner if in fact he will lose? Suppose that he will receive 30% of the votes $(p = 0.3)$. What is the probability β that the sample will erroneously lead us to conclude that H_0 is true and that Jones is going to win?

Example

SOLUTION 10.2

$$\beta = P(type\ II\ error) = P(accepting\ H_0\ when\ Ha\ is\ true)$$
$$= P(value\ of\ the\ test\ statistic\ is\ not\ in\ RR\ when\ H_a\ is\ true).$$

Because we want to calculate β when $p = 0.3$ (a particular value of p that is in H_a),

$$\beta = P(Y > 2\ when\ p = 0.3) = \Sigma_{y=3}^{15} \binom{15}{y} (0.3)^y (0.7)^{15-y}.$$

Again consulting Table 1, Appendix 3, we find that $\beta = .873$. If we use $RR = \{y \leq 2\}$, our test will usually lead us to conclude that Jones is a winner (with probability $\beta = .873$), even if p is as low as $p = 0.3$.

Example

EXAMPLE 10.3

Refer to Example 10.1. and 10.2. Calculate the value of β if Jones will receive only 10% of the votes ($p = .1$)

Example

SOLUTION 10.3

$$\beta = P(type\ II\ error) = P(accepting\ H_0\ when\ p = .1)$$
$$= \ P(value\ of\ the\ test\ statistic\ is\ not\ in\ RR\ when\ p = .1)$$
$$= P(Y > 2\ when\ p = 0.1) = \Sigma_{y=3}^{15} \binom{15}{y} (0.1)^y (0.9)^{15-y} = .184.$$

Example

EXAMPLE 10.4

Refer to Example 10.1. Now assume that $RR = \{y \leq 5\}$. Calculate the level α of the test and calculate β if $p = .3$. Compare the results with the values obtained in Examples 10.1 and 10.2 (where we used $RR = \{y \leq 2\}$.

Example

SOLUTION 10.4

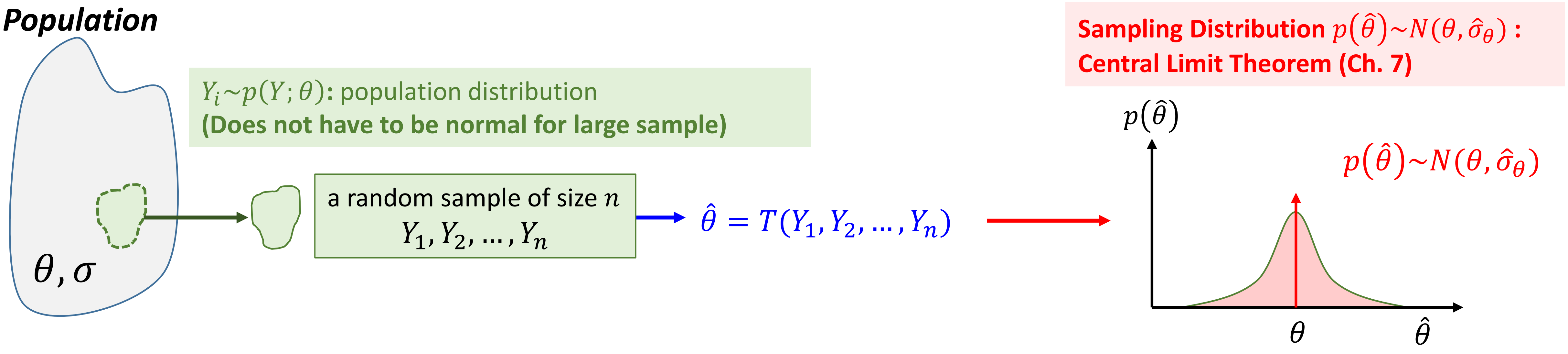
$$\alpha = P(\textit{test statistic is in RR when } H_0 \textit{ is true}) = P(Y \le 5 | p = .5) = \sum_{y=6}^{15} \binom{15}{y} (.5)^{15} = .151.$$

$$\beta = P(\textit{test statistic is not in RR when } H_a \textit{ is true}) = P(Y > 5 | p = .3) = \sum_{y=6}^{15} \binom{15}{y} (.3)^{y} (.7)^{15-y}$$

$$= .278.$$

Note that enlarging the rejection region increased α and decreased β.

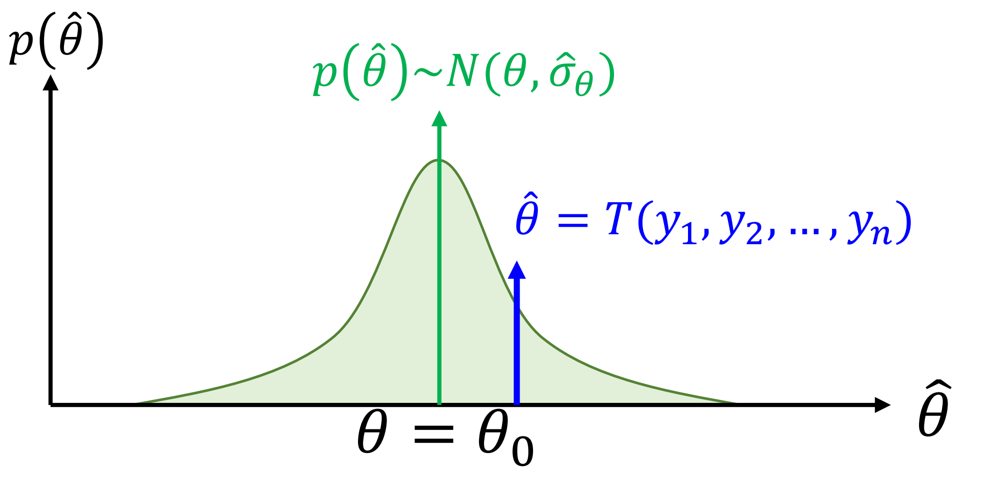
Motivation



- In this section, we will develop hypothesis-testing procedures that are based on an estimator $\hat{\theta}$ that has an (approximately) normal sampling distribution $p(\hat{\theta}) \sim N(\theta, \hat{\sigma}_\theta)$ with mean θ and standard error $\hat{\sigma}_\theta$
- The large-sample estimators of Chapter 8 (Table 8.1), such as $\bar{Y}$ and $\hat{p}$, satisfy these requirements.
- So do the estimators used to compare of two population means $(\mu_1 - \mu_2)$ and for the comparison of two binomial parameters $(p1 - p2)$.

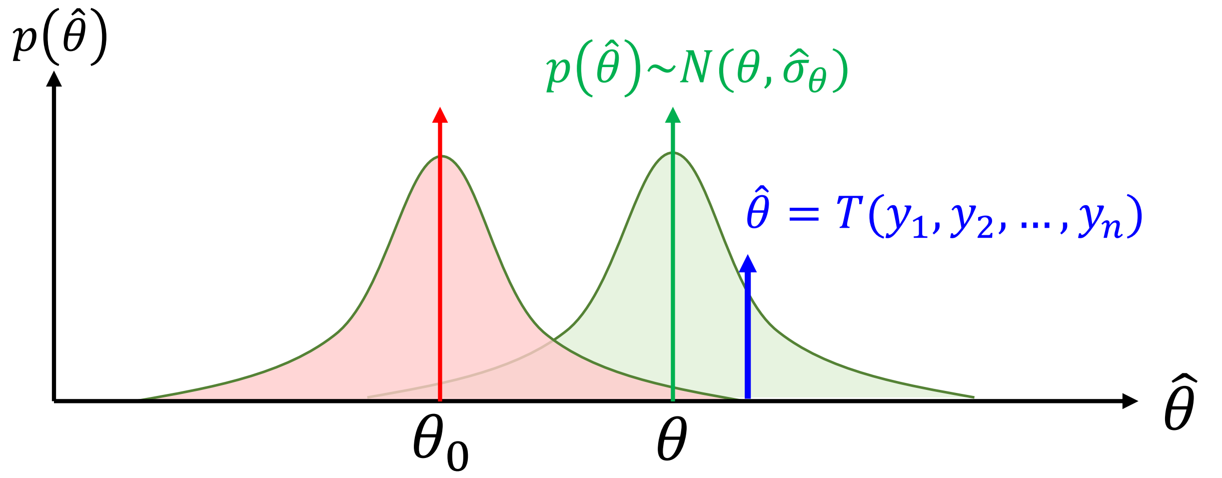
Procedure

- When H_0: $\theta = \theta_0$ is true:



Test statistic of a realized sample $\hat{\theta} = T(y_1, y_2, \dots, y_n)$ will be close to θ_0 ⇒ **Accept** H_0: $\theta = \theta_0$

- When H_a: $\theta > \theta_0$ is true:



Test statistic of a realized sample $\hat{\theta} = T(y_1, y_2, \dots, y_n)$ will be larger than θ_0 ⇒**Reject** H_0: $\theta = \theta_0$

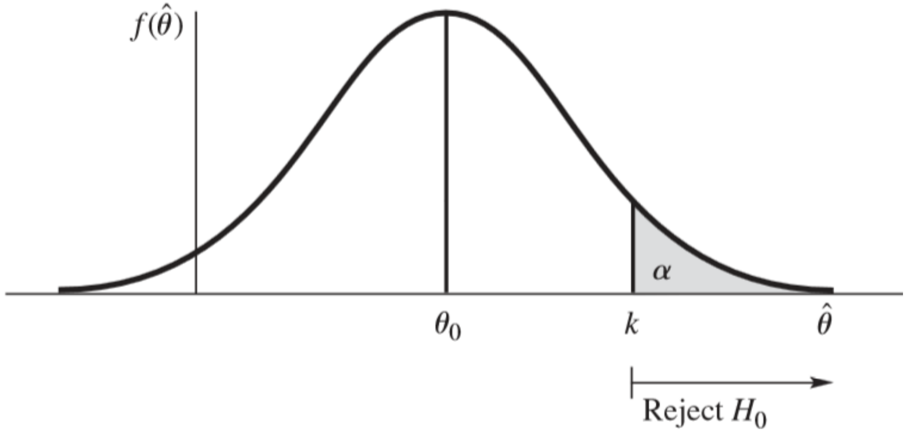
Procedure

- The null and alternative hypotheses, the test statistic, and the rejection region are as follows:

  - ✓ $H_0: \theta = \theta_0$
  - ✓ $H_a: \theta > \theta_0$
  - ✓ Test statistic: $\hat{\theta}$
  - ✓ Rejection region: RR= $\{\hat{\theta} > k\}$ for some choice of k

Choosing Rejection Region

- The actual value of k in the rejection region RR is determined by fixing the type I error probability α (the level of the test) and choosing k accordingly.
- If H_0 is true, $\hat{\theta}$ has an approximately normal distribution $p(\hat{\theta}) \sim N(\theta_0, \hat{\sigma}_\theta)$ with mean θ_0 and standard error $\hat{\sigma}_\theta$



α –Level test

$$P(\textit{type I error}) = \alpha$$

$$P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$$

$$P(\hat{\theta} \in RR | \theta = \theta_0) = \alpha$$

$$P(\hat{\theta} > k | \theta = \theta_0) = \alpha$$

$$k = \theta_0 + z_\alpha \hat{\sigma}_\theta$$

$$RR = \{\hat{\theta}: \hat{\theta} > \theta_0 + z_\alpha \hat{\sigma}_\theta\} = \left\{\hat{\theta}: \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta} > z_\alpha\right\}$$

If $Z = (\hat{\theta} - \theta_0)/\hat{\sigma}_\theta$ is used as test statistics, the rejection region can be written as $RR = \{z > z_\alpha\}$

Procedure

- The null and alternative hypotheses, the test statistic, and the rejection region for general test

  ✓ $H_0: \theta = \theta_0$
  ✓ $H_a: \theta > \theta_0$
  ✓ Test statistic: $\hat{\theta}$
  ✓ Rejection region: RR= $\{\hat{\theta} > k\}$ for some choice of k

- Large-Sample test

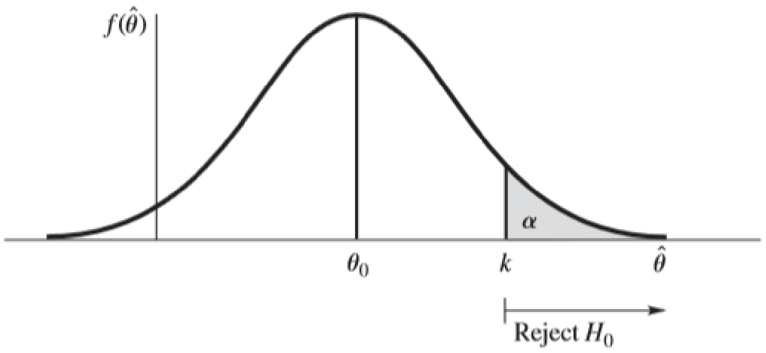
  ✓ $H_0: \theta = \theta_0$
  ✓ $H_a: \theta > \theta_0$ (upper tail alternative)
  ✓ Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
  ✓ Rejection region:RR= $\{z > z_\alpha\}$: (upper tail rejection region)

$$Z = \frac{\text{estimator for the parameter} - \text{value for the parameter givey by } H_0}{\text{standard error of the esitimator}}$$

- H_0 is rejected if Z falls far enough into the upper tail of the standard normal distribution

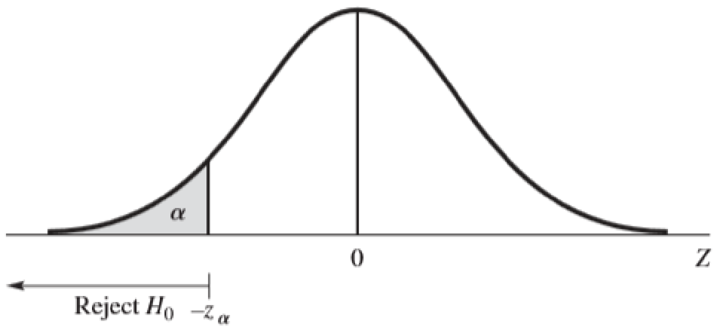
Upper, Lower, Two-tailed Hypothesis Tests

Testing $H_0: \theta = \theta_0$ against $H_0: \theta > \theta_0$



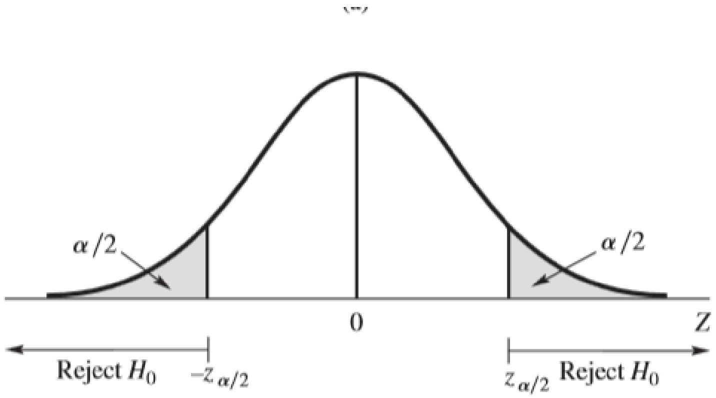
- $H_0: \theta = \theta_0$
- $H_a: \theta > \theta_0$ (upper tail alternative)
- Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- Rejection region:RR= $\{z > z_\alpha\}$ (upper tail rejection region)

Testing $H_0: \theta = \theta_0$ against $H_0: \theta < \theta_0$



- $H_0: \theta = \theta_0$
- $H_a: \theta < \theta_0$ (lower tail alternative)
- Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- Rejection region:RR= $\{z < -z_\alpha\}$ (lower tail rejection region)

Testing $H_0: \theta = \theta_0$ against $H_0: \theta \neq \theta_0$



- $H_0: \theta = \theta_0$
- $H_a: \theta \neq \theta_0$ (two-sided alternative)
- Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- Rejection region:RR= $\left\{|z\} > z_{\alpha/2}\right\}$: (two-sided rejection region)

How do we decide which alternative to use for a test? ➔ Depends on the hypothesis that we seek to support.

Example

EXAMPLE 10.5

A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.) As a check on his claim, $n = 36$ salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. The mean and variance of the 36 measurements were 17 and 9, respectively. Does the evidence contradict the vice president's claim? Use a test with level $\alpha = 0.05$.

Example

SOLUTION 10.5

$H_0 : \mu = 15$ against $H_a : \mu > 15$.

We know that for large enough n, the sample mean $\bar{Y}$ is a point estimator of μ that is approximately normally distributed with $\mu_{\bar{Y}} = \mu$ and $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$. Hence, our test statistic is $Z = \frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}}$.

The rejection region, with $\alpha = 0.05$, is given by $\{z > z_{0.05} = 1.645\}$ (see Table 4, Appendix 3).

The population variance σ^2 is not known, but it can be estimated very accurately (because n = 36 is sufficiently large) by the sample variance $s^2 = 9$.

Example

EXAMPLE 10.6

A machine in a factory must be repaired if it produces more than 10% defectives among the large lot of items that it produces in a day. A random sample of 100 items from the day's production contains 15 defectives, and the supervisor says that the machine must be repaired. Does the sample evidence support his decision? Use a test with level .01.

Example

SOLUTION 10.6

$H_0 : p = .10$ against $H_a : p > .10$.

Test statistic $Z = \frac{\hat{p}-p_0}{\sigma_{\hat{p}}} = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ where $\hat{p} = Y/n$

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{.15 - .10}{\sqrt{(.1)(.9)/100}} = \frac{5}{3} = 1.667.$$

Since $P(Z > 2.33) = .01$, the observed test statistic is not in the rejection region, we cannot reject the null hypothesis, and the evidence does not support the supervisor's decision.

Example

EXAMPLE 10.7

A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are shown in Table 10.2. Do the data present sufficient evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = 0.05$.

Table 10.2 Data for Example 10.7

Men	Women
$n_1 = 50$	$n_2 = 50$
$\overline{y}_1 = 3.6$ seconds	$\overline{y}_2 = 3.8$ seconds
$s_1^2 = .18$	$s_2^2 = .14$

Example

SOLUTION 10.7

$H_0 : \mu_1 - \mu_2 = 0$ against $H_a : \mu_1 - \mu_2 \neq 0$ where μ_1 and μ_2 denote the true mean reaction times for men and women, respectively.

Test statistic $Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, where σ_1^2, σ_2^2 are respective population variances.

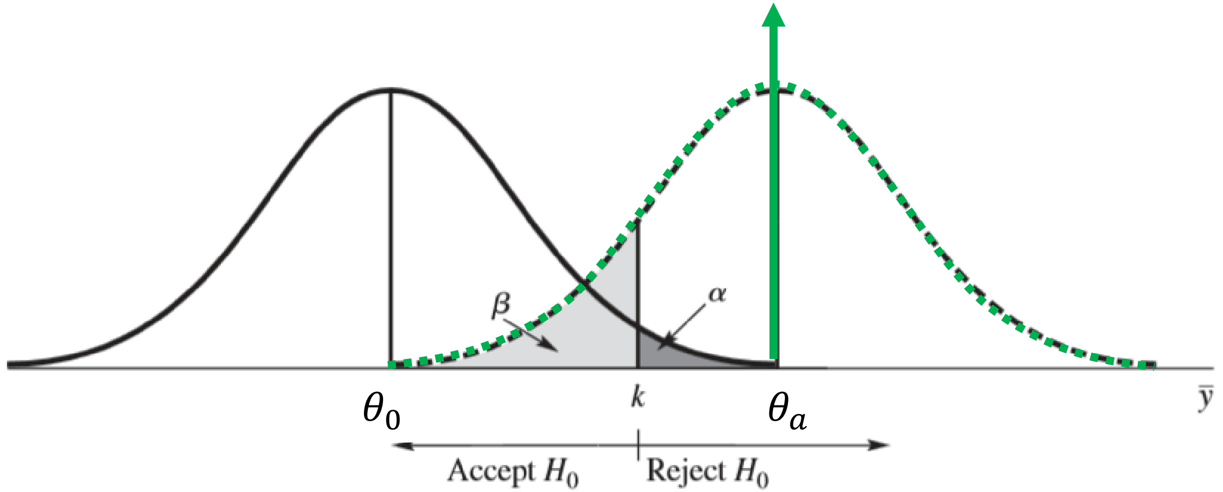
For large sample, sample variances are good estimates of their corresponding population variances, so $z \simeq \frac{3.6 - 3.8}{\sqrt{\frac{.18}{50} + \frac{.14}{50}}} = -2.5.$

Since $P(|z| > 1.96) = 0.05$, the value falls in the rejection region, and we conclude that mean reaction times differ for men and women.

Motivation

- For the test $H_0: \theta = \theta_0$ versus $H_a: \theta > \theta_0$, we can calculate *type II error* probabilities β only for specific values for θ in H_a.
- Suppose that the experimenter has in mind a specific alternative—say, $\theta = \theta_a$(where $\theta_a > \theta_0$).
- Because the rejection region is of the form $\mathrm{RR} = \{\hat{\theta}: \hat{\theta} > k\}$
- The probability β of a type II error is

$$\begin{aligned}\beta &= P(\hat{\theta} \text{ is not in RR} \mid H_a \text{ is true}) \\ &= P(\hat{\theta} \leq k \mid \theta = \theta_a) \\ &= P\left(\frac{\hat{\theta} - \theta_a}{\hat{\sigma}_\theta} \leq \frac{k - \theta_a}{\hat{\sigma}_\theta} \middle| \theta = \theta_a\right)\end{aligned}$$



- For a fixed sample of size n, the size of β depends on the distance between θ_a and θ_0.
  - If θ_a is close to θ_0, the true value of θ (either θ_0 or θ_a) is difficult to detect,
  - the probability of accepting H_0 when H_a is true tends to be large.

Example

EXAMPLE 10.8

Suppose that the vice president in Example 10.5 wants to be able to detect a difference equal to one call in the mean number of customer calls per week. That is, he wishes to test $H_0 : \mu = 15$ against $H_a : \mu = 16$. With the data as given in Example 10.5, find β for this test.

Example

SOLUTION 10.8

The rejection region for a .05 level test was given by $z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} \simeq \frac{\bar{y}-\mu_0}{s/\sqrt{n}} = \frac{\bar{y}-15}{3/\sqrt{36}} > 1.645$ or $\bar{y} > 15.8225$.

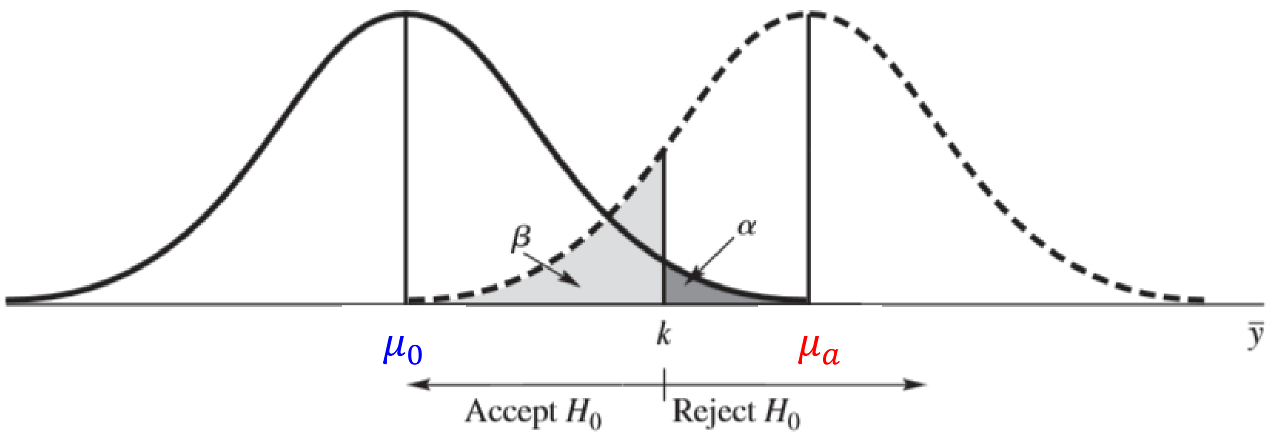
$$\beta = P\left(\frac{\bar{Y}-\mu_a}{\sigma/\sqrt{n}} \leq \frac{15.8225-16}{3/\sqrt{36}}\right) = P(Z \leq -.36) = .3594.$$

Selecting Sample Size

- Suppose that you want to test $H0 : \mu = \mu_0$versus $H_a : \mu > \mu_0$.
- If you specify the desired values of α and β (where β is evaluated when $\mu = \mu_a$ and $\mu_a > \mu_0$), any further adjustment of the test must involve two remaining quantities:
  - ✓ The sample size n
  - ✓ The point at which the rejection region begins, k.
- Because α and β can be written as probabilities involving n and k, we have two equations in two unknowns, which can be solved simultaneously for n. Thus,

$$\alpha = P(\bar{Y} > k|\mu = \mu_0)$$
$$= P\left(\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}} \middle| \mu = \mu_0\right) = P(Z > z_\alpha)$$

$$\beta = P(\bar{Y} \leq k|\mu = \mu_a)$$
$$= P\left(\frac{\bar{Y} - \mu_a}{\sigma/\sqrt{n}} \leq \frac{k - \mu_a}{\sigma/\sqrt{n}} \middle| \mu = \mu_a\right) = P(Z < -z_\beta)$$



Selecting Sample Size

$$\alpha = P(\bar{Y} > k|\mu = \mu_0)$$
$$= P\left(\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}} > \frac{k-\mu_0}{\sigma/\sqrt{n}} \middle| \mu = \mu_0\right) = P(Z > z_\alpha) \qquad \Rightarrow \frac{k-\mu_0}{\sigma/\sqrt{n}} = z_\alpha \qquad \text{-------------(1)}$$

$$\beta = P(\bar{Y} \le k|\mu = \mu_a)$$
$$= P\left(\frac{\bar{Y}-\mu_a}{\sigma/\sqrt{n}} \le \frac{k-\mu_a}{\sigma/\sqrt{n}} \middle| \mu = \mu_a\right) = P(Z < -z_\beta) \qquad \Rightarrow \frac{k-\mu_0}{\sigma/\sqrt{n}} = -z_\beta \qquad \text{-------------(2)}$$

Solving both of the above equations for k gives

$$k = \mu_0 + z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - z_\beta\left(\frac{\sigma}{\sqrt{n}}\right)$$

Thus,

$$(z_\alpha + z_\beta)\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - \mu_0, \qquad \text{or equivently} \quad n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2}$$

Example

EXAMPLE 10.9

Suppose that the vice president of Example 10.5 wants to test $H_0 : \mu = 15$ against $H_a : \mu = 16$ with $\alpha = \beta = .05$. Find the sample size that will ensure this accuracy. Assume that σ^2 is approximately 9.

Example

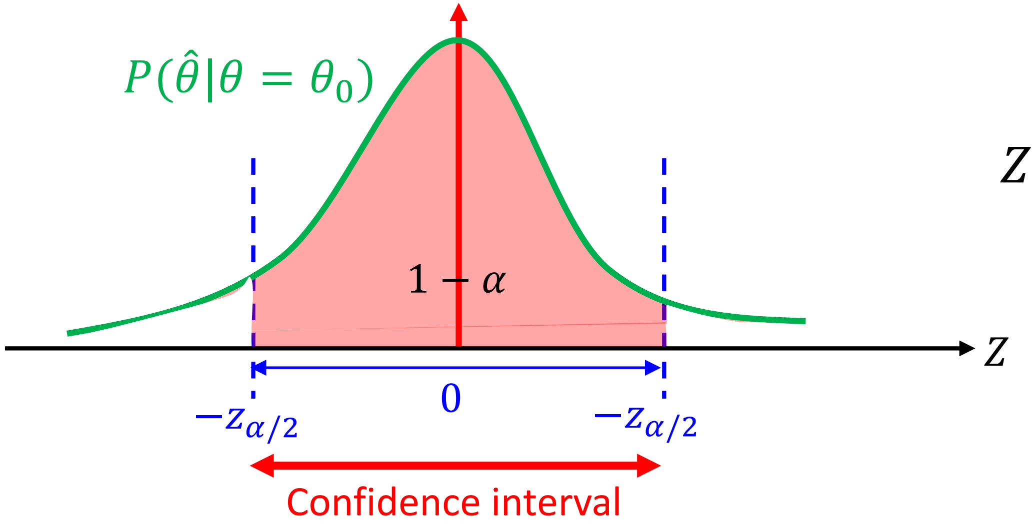
SOLUTION 10.9

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{(\mu_a - \mu_0)^2} = \frac{(1.645 + 1.645)^2 (9)}{(16 - 15)^2} = 97.4.$$

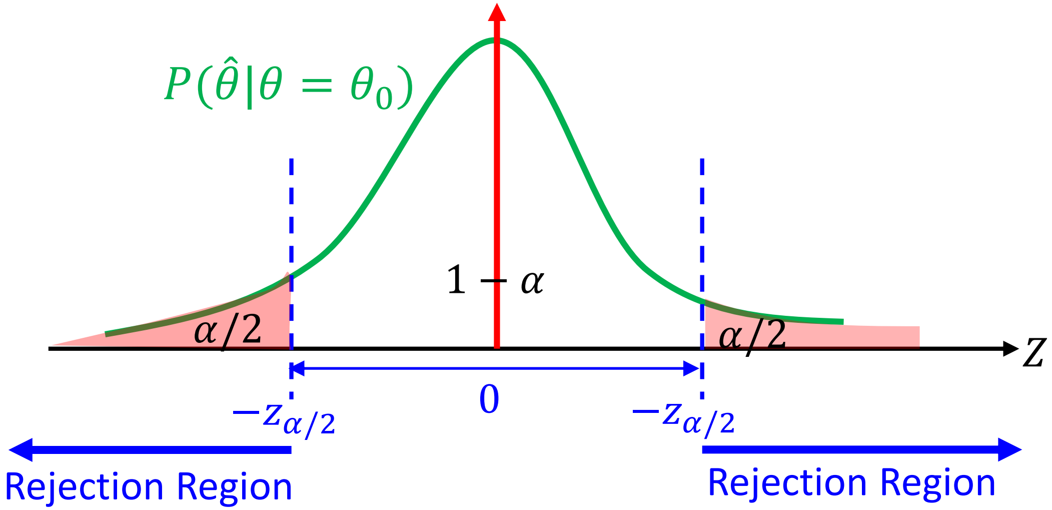
Thus, $n = 98$ observations should be used to meet the requirements.

Motivation

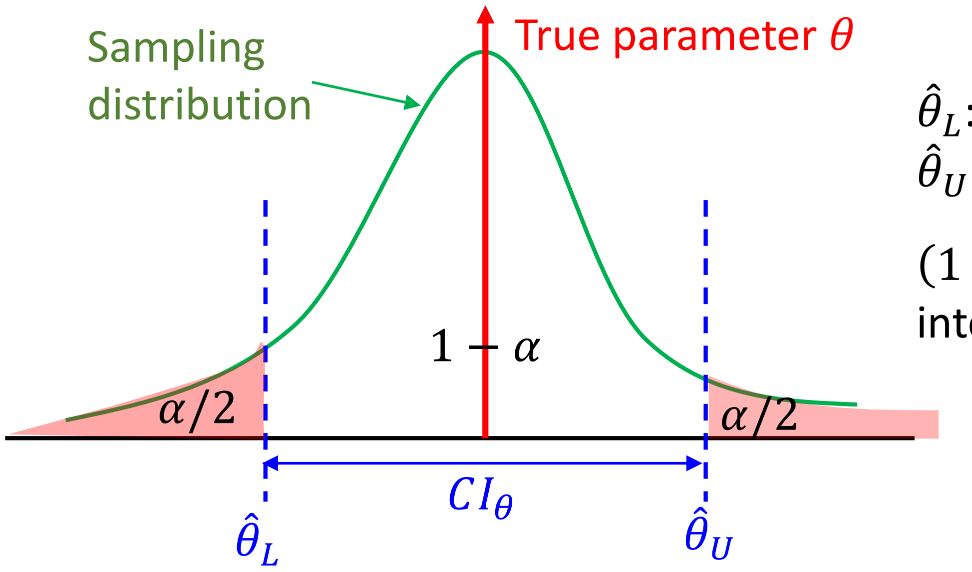
- What is Relationships Between Hypothesis-Testing Procedure?



$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$



Recall: Confidence Interval



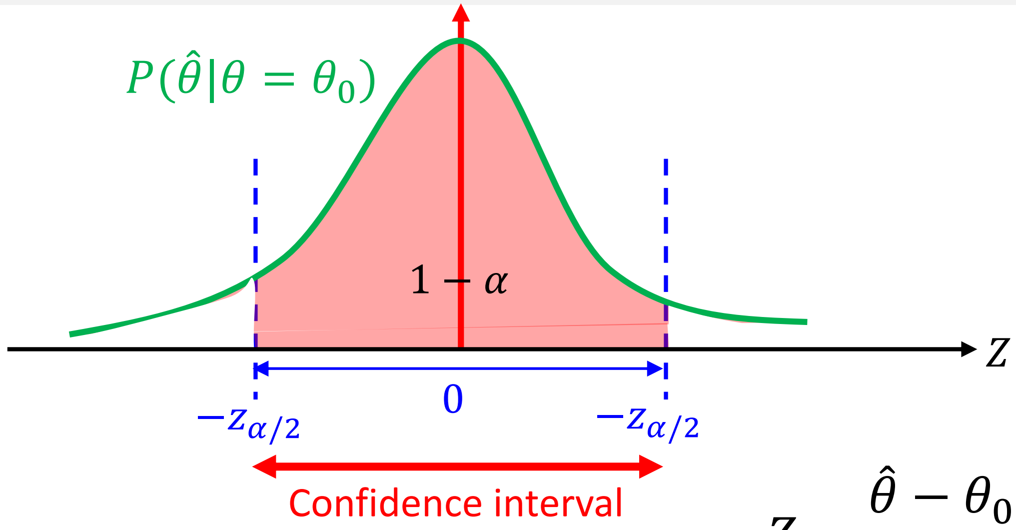
$\hat{\theta}_L$: *The lower confidence limit*, which is a random (function of a random samples)
$\hat{\theta}_U$: *The upper confidence limit*, which is a random (function of a random samples)

$(1-\alpha)$: confidence coefficient, the probability that a (random) confidence interval will enclose θ (a fixed quantity) is called the confidence coefficient

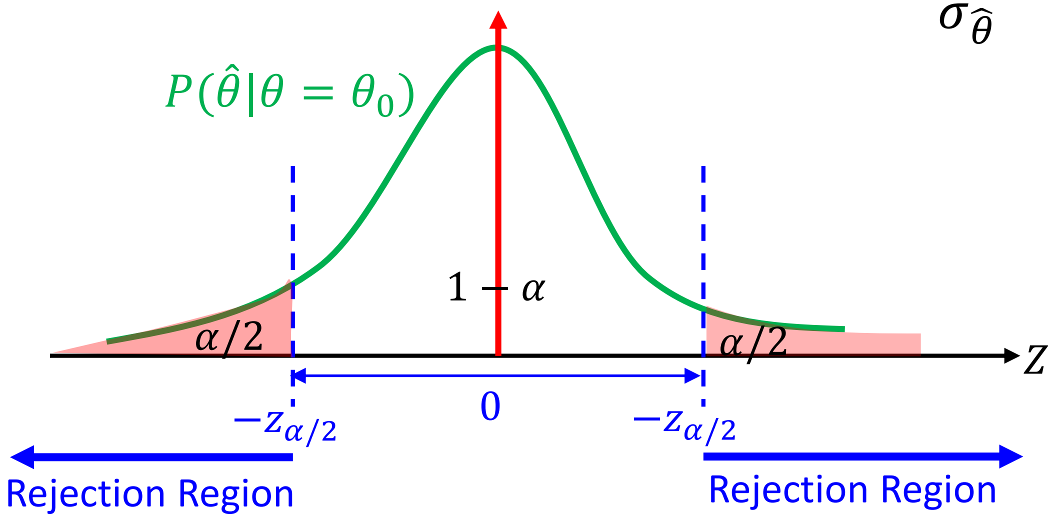
$$P\left(\hat{\theta}_L \le \theta \le \hat{\theta}_L\right) = 1-\alpha$$

- *"There is a $(1-\alpha)$ % probability that when I compute the confidence interval (CI) from a current data sample, the computed CI contains θ*
  → From current data set, We can only say that $\theta \in \mathrm{CI}$ or $\theta \notin \mathrm{CI}$
- From a practical point of view, the confidence coefficient identifies the fraction of the time, **in repeated sampling**, that the intervals constructed will contain the target parameter θ.
  - If the confidence coefficient is high, we can be highly confident that any confidence interval, **constructed by using the results from a single sample**, will enclose θ.

Confidence Interval vs. Rejection Region



$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$



$100(1-\alpha)\%$ Confidence interval

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

$$P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta_0 \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$$

The null hypothesis H_0 is **not rejected** (is accepted) at level α
In the $100(1-\alpha)\%$ Confidence interval

$\alpha -$ level Rejection region:RR= $\{|z\} > z_{\alpha/2}\}$

$$P(Z < -z_{\alpha/2}, z_{\alpha/2} < Z) = \alpha$$

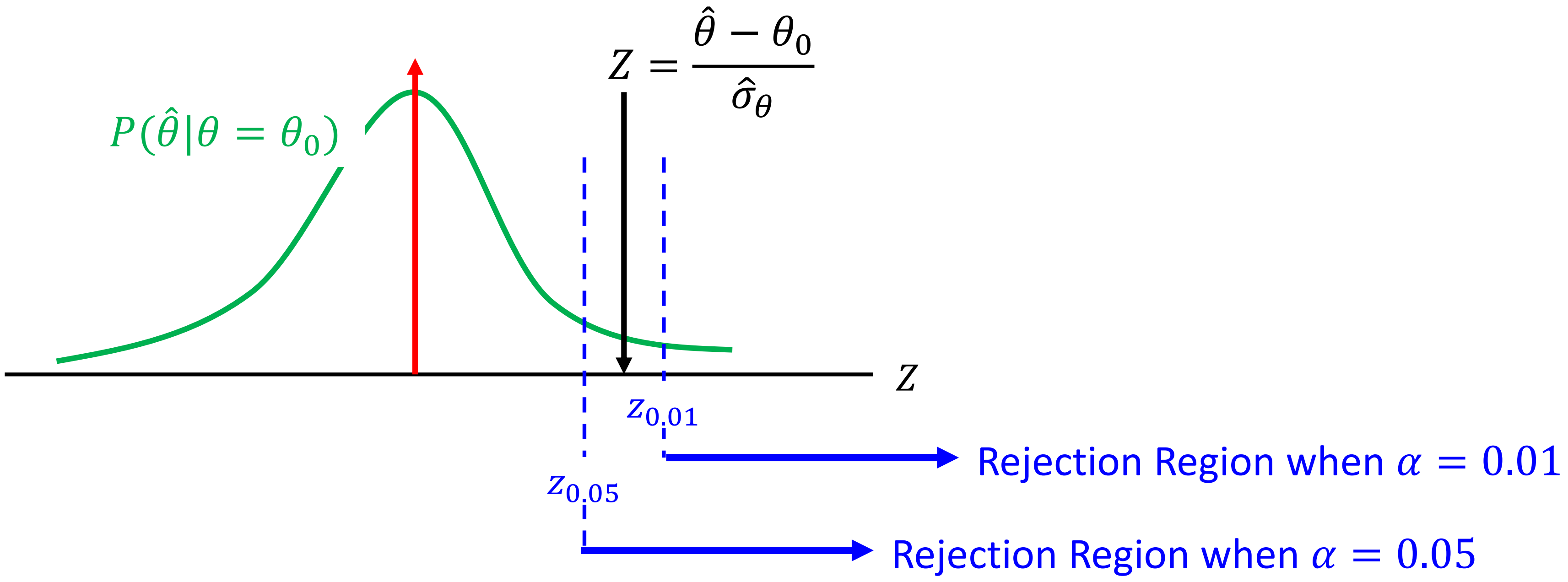
$$P(\theta_0 < \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}} < \theta_0) = \alpha$$

The null hypothesis H_0 is **rejected** at level α

- Thus, a duality exists between our large-sample procedures for constructing a $100(1-\ \alpha)\%$ two-sided confidence interval and for implementing a two-sided hypothesis test with level α.

Motivation

- the probability α of a type I error is often called the significance level, or, more simply, the level of the test

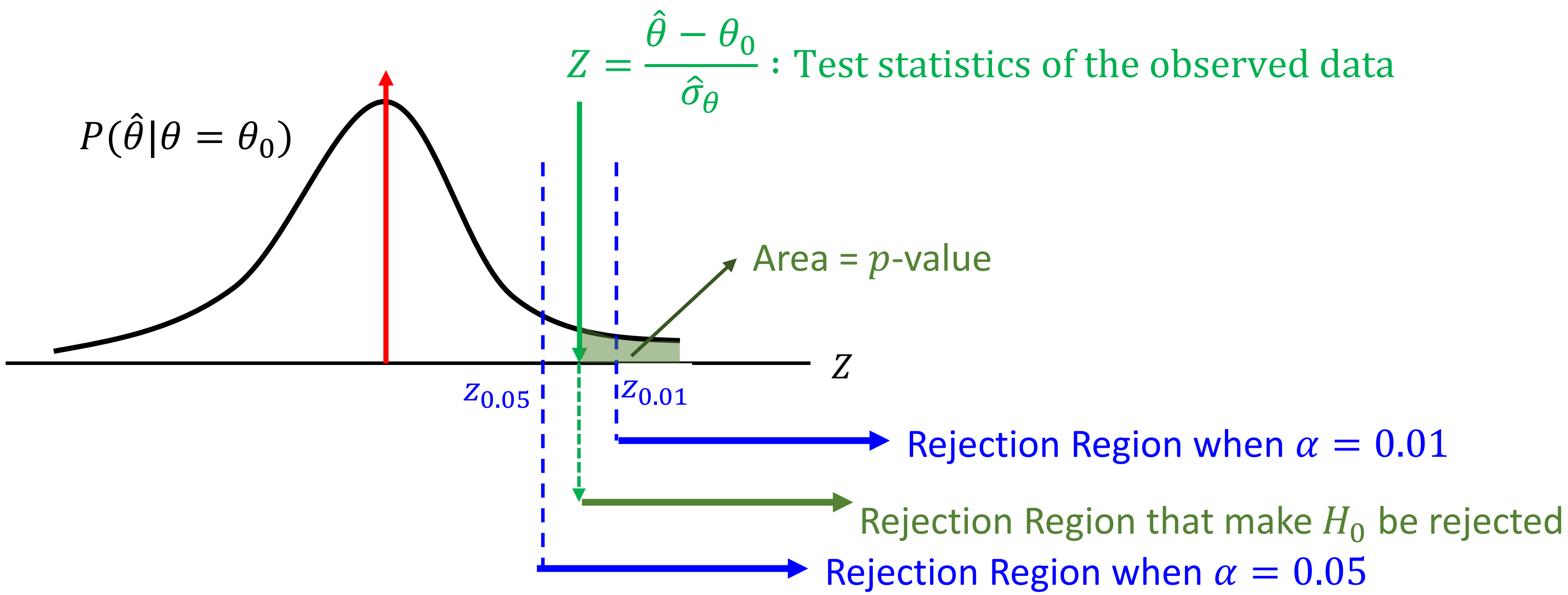


- It is possible, therefore, for two persons to analyze the same data and reach opposite conclusions
  - ✓ one concluding that the null hypothesis should ***be rejected*** at the α = .05 significance level
  - ✓ the other deciding that the null hypothesis should ***not be rejected*** with α = .01.
- Although small values of α are often recommended, the actual value of α to use in an analysis is somewhat arbitrary.
  - ✓ often are used out of habit or for the sake of convenience

p-value

DEFINITION 10.2

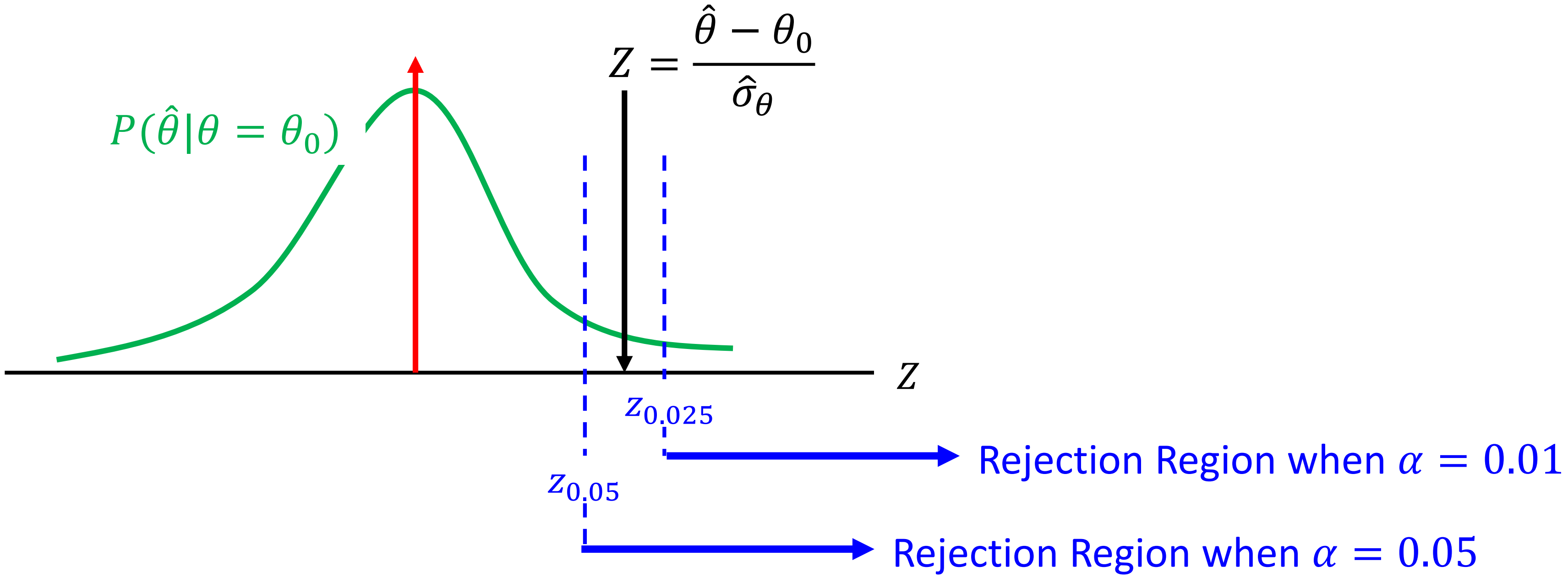
If *W* is a test statistic, the *p-value*, or *attained significance level*, is **the smallest level** of significance α for which the observed data indicate that the null hypothesis should be rejected.



- If $\alpha \geq p - \text{value}$, H_0 will be always rejected
- If $\alpha < p - \text{value}$, H_0 will not be rejected

➢ the p-value allows the reader of papers to evaluate *the extent* to which the observed data disagree with H_0

Definition



- If a test result is statistically significant for $\alpha = .05$ but not for $\alpha = .025$, we will report

$$.025 \leq p - value \leq .05.$$

- Thus, for any $\alpha \geq .05$, we reject the null hypothesis;
- For $\alpha < .025$, we do not reject the null hypothesis;
- For values of α that fall between .025 and .05, we need to seek more complete tables of the appropriate distribution before reaching a conclusion. The tables in the appendix provide useful information about p-values, but the results are usually rather cumbersome.

Example

EXAMPLE 10.10

Recall our discussion of the political poll (see Examples 10.1 through 10.4) where $n = 15$ voters were sampled. If we wish to test $H_0\colon\ p = .5$ versus $H_a\colon\ p < .5$, using $Y =$ the number of voters favoring Jones as our test statistic, what is the p-value if $Y = 3$? Interpret the result.

Example

SOLUTION 10.10

H_0 is rejected for small values of Y. Thus, the p-value for this test is $P(Y \le 3)$, where Y has a binomial distribution $n = 15$ and $p = .5$. Using Table 1, Appendix 3, we find that the p-value is .018. Because the p-value is the smallest α for which the null hypothesis is rejected, we conclude that Jones does not have a plurality of the vote for $\alpha \ge .018$, while the null hypothesis could not be rejected for $\alpha \le .018$.

Example

EXAMPLE 10.11

Find the p-value for the statistical test of Example 10.7.

Example

SOLUTION 10.11

Example 10.7 presents a test of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$. The value of the computed test statistic was $z = -2.5$. Because this test is two-tailed, the p-value is

$$P(|Z| \geq 2.5) = P(Z \geq 2.5 \; or \; Z \leq -2.5) = 2(.0062) = .0124.$$

Comments

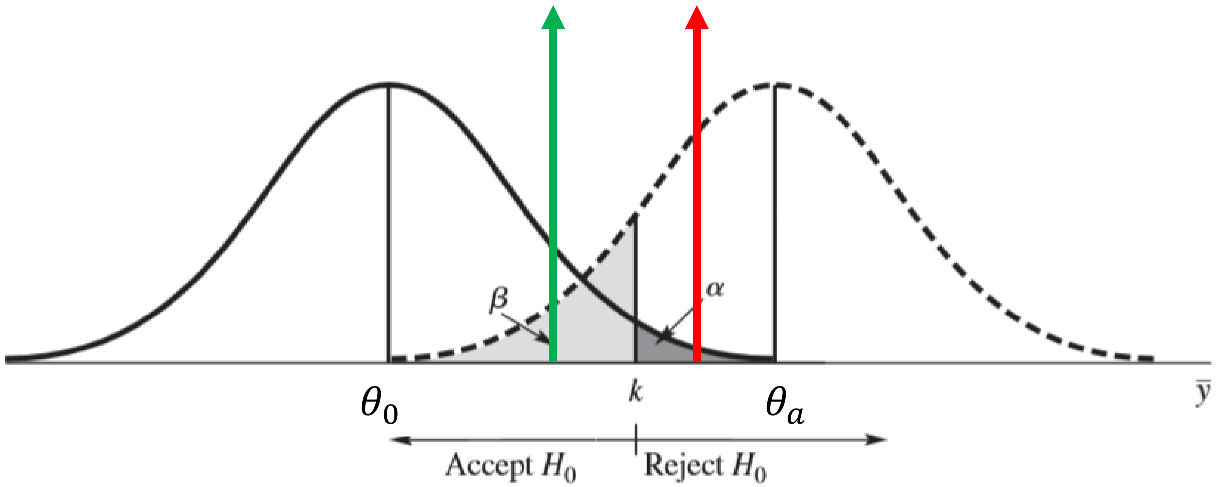
- We can choose between implementing a one-tailed or a two tailed test for a given situation.

- The probability β of a *type II error* can be calculated only after a specific value of the parameter of interest has been singled out for consideration.
  - ✓ The selection of a practically meaningful value for this parameter is often difficult

- Later in this chapter, we will determine methods for selecting tests with the smallest possible value of β for tests where α, the probability of a type I error, is a fixed value selected by the researcher.

Comments

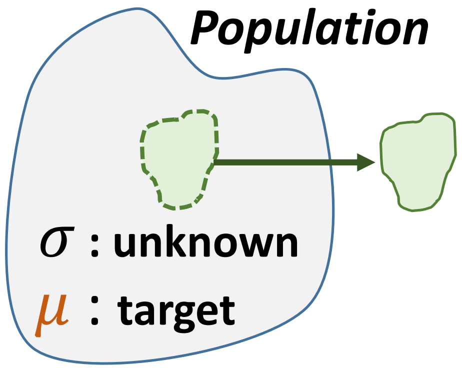
- When the **value of the test statistic** is not in the rejection region, we will "**fail to reject**" rather than "accept" the null hypothesis.
  - If, however, Y **does not fall in the rejection region** and we can determine no specific value of p in H_a that is of direct interest, we simply state that we will not reject H_0 and must seek additional information before reaching a conclusion. (Decision is only regarding H_0 not in H_a)



	H_0 is True	H_0 is False
Reject H_0	*type I error* $P(\textit{type I error}) = \alpha$	Correct Decision
Accept H_0	Correct Decision	*type II error* $P(\textit{type II error}) = \beta$

- If $\boldsymbol{H_0}$ **is rejected** for a "small" value of α (or for a small p-value), this occurrence does not imply that the null hypothesis is "wrong by a large amount."
  - ➢ It does meant hat the null hypothesis can be rejected based on a procedure that incorrectly rejects the null hypothesis (when H_0 is true) with a small probability (that is, with a small probability of a *type I error*).

Motivation



a random sample of size n
$Y_1, Y_2, \dots, Y_n$

Sample mean

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

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Sample variance

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2, df = (n-1)$$

$$\frac{Z}{\sqrt{W/v}} = \sqrt{n}\left(\frac{\bar{Y}-\mu}{S}\right) \sim t, df = (n-1)$$

- When $Y_i \sim N(\mu, \sigma^2)$
- When n is **small**
- When σ is assumed to be **unknown**

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{S}\right) \sim t, df = (n-1)$$

We develop formal procedures for testing hypotheses about μ and $\mu_1 - \mu_2$, procedures that are appropriate for **small samples** from **normal populations**.

Small-Sample Test for μ

- We assume that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a normal distribution with unknown mean μ and unknown variance σ^2. If Y and S denote the sample mean and sample standard deviation, respectively, and if $H_0 : \mu = \mu_0$ is true, then

$$T = \sqrt{n}\left(\frac{\bar{Y} - \mu_0}{S}\right)$$

has a t distribution with $n - 1$ df

- Because the t distribution is symmetric and mound-shaped, the rejection region for a small-sample test of the hypothesis $H_0 : \mu = \mu_0$must be located in the tails of the t distribution and be determined in a manner similar to that used with the large-sample Z statistic.

Example

EXAMPLE 10.12

Example 8.11 gives muzzle velocities of eight shells tested with a new gunpowder, along with the sample mean and sample standard deviation, $\bar{y} = 2959$ and $s = 39.1$. The manufacturer claims that the new gunpowder produces an average velocity of not less than 3000 feet per second. Do the sample data provide sufficient evidence to contradict the manufacturer's claim at the .025 level of significance?

Example

SOLUTION 10.12

Assuming that muzzle velocities are approximately normally distributed, we want to test $H_0: \mu = 3000$ versus $H_a: \mu < 3000$.

The rejection region is given by $t < -t_{0.025} = -2.365$, where t possesses $(n-1) = 7$ df.

The observed value of test statistic is

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{2959 - 3000}{39.1/\sqrt{8}} = -2.966.$$

This value falls in the rejection region, and the null hypothesis is rejected at the $\alpha = .025$ level of significance.

Example

EXAMPLE 10.13

What is the p-value associated with the statistical test in Example 10.12?

Example

SOLUTION 10.13

Because the null hypothesis should be rejected if t is "small", p-value = $P(T < -2.966)$, where T has a t distribution with $n - 1 = 7$ df.

Since it is tiresome to compute the exact value, we may impose bounds on the p-value. Table 5 in Appendix 3 shows that $-t_{.025} = -2.365$ and $-t_{.01} = -2.998$ thanks to the symmetry of t distribution. Thus, we conclude that $.01 \leq p - \text{value} \leq .025$.

Example

EXAMPLE 10.14

Example 8.12 gives data on the length of time required to complete an assembly procedure using each of two different training methods. The sample data are as shown in Table 10.3. Is there sufficient evidence to indicate a difference in true mean assembly times for those trained using the two methods? Test at the $\alpha = .05$ level of significance.

Table 10.3 Data for Example 10.14

Standard Procedure	New Procedure
$n_1 = 9$	$n_2 = 9$
$\overline{y}_1 = 35.22$ seconds	$\overline{y}_2 = 31.56$ seconds
$\sum_{i=1}^{9}(y_{1i} - \overline{y}_1)^2 = 195.56$	$\sum_{i=1}^{9}(y_{2i} - \overline{y}_2)^2 = 160.22$

Example

SOLUTION 10.14

We are testing $H_0: \mu_1 - \mu_2 = 0$ against $H_a: \mu_1 - \mu_2 \neq 0$. Consequently, we must muse a two-tailed test.

The test statistic is $T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with $\mu_1 - \mu_2 = 0$, and the rejection region for $\alpha = .05$ is $|t| > t_{.025} = 2.120$, since t is based on $(n_1 + n_2 - 2) = 9 + 9 - 2 = 16$ df.

Note that $s_p = \sqrt{s_p^2} = \sqrt{\frac{195.56 + 160.22}{9+9-2}} = \sqrt{22.24} = 4.716.$

Then, $t = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{35.22 - 31.56}{4.715\sqrt{\frac{1}{9} + \frac{1}{9}}} = 1.65.$

This value does not fall in the rejection region, hence, the null hypothesis is not rejected.

Example

EXAMPLE 10.15

Find the p-value for the statistical test in Example 10.14.

Example

SOLUTION 10.15

The p-value for this test is $P(T > 1.65 \; or \; T < -1.65)$.
Because this test statistic is based on $n_1 + n_2 - 2 = 16$ df, we consult Table 5, Appendix 3, to find $t_{0.05} = 1.746$ and $t_{0.10} = 1.337$. Thus, $0.05 < P(T > 1.65) < 0.10$, and $0.05 < P(T < -1.65) < 0.10$. We conclude that $0.10 < p - \text{value} < 0.20$.