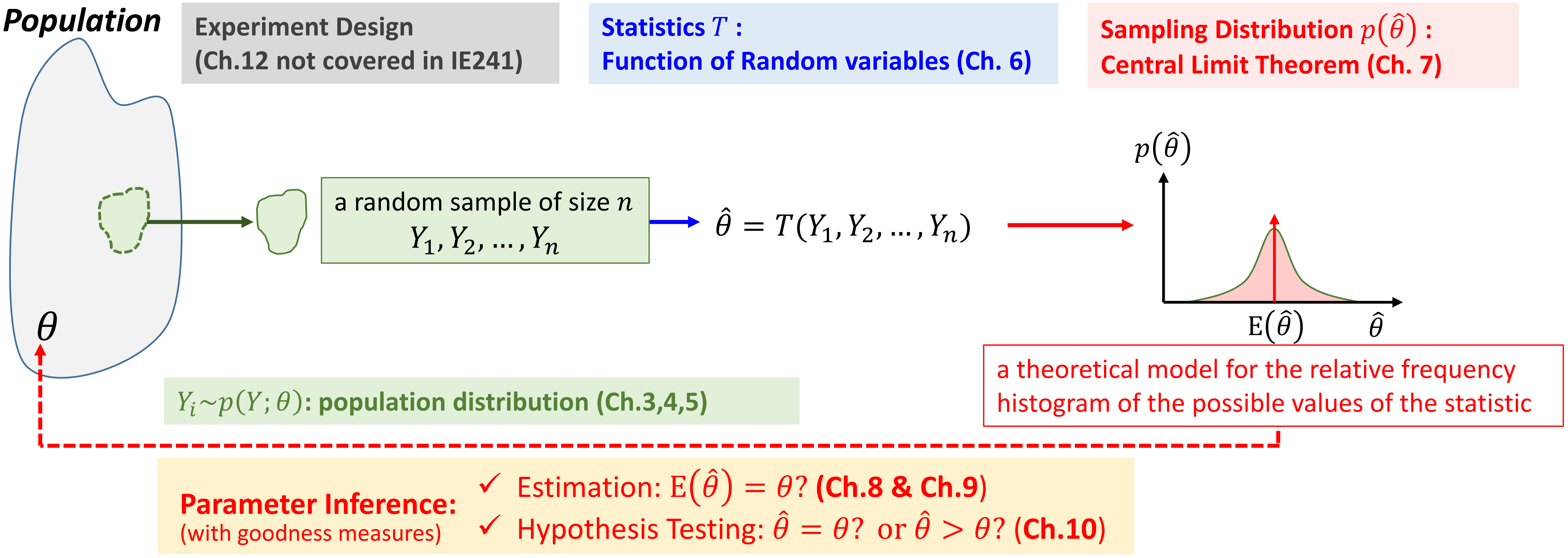
CHAPTER 10
Hypothesis Testing

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.

Make a hypothesis regarding nature → Conduct experiments (observations) → Accept or Reject

Make a hypothesis regarding population parameters → Draw Samples (observations) → Accept or Reject

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

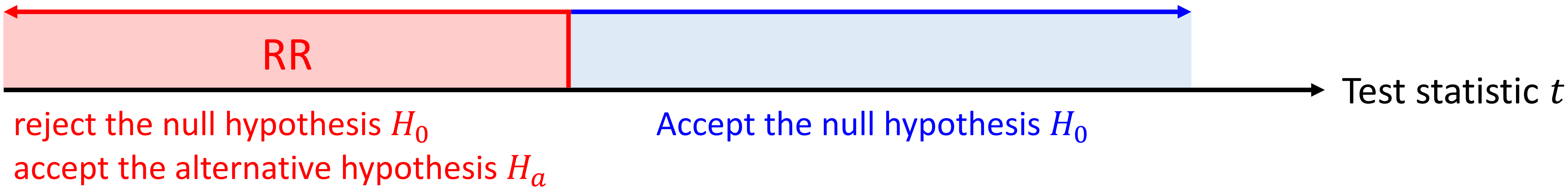
What would you like to support? → Alternative hypothesis, H_a

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 $\mathrm{RR} = \{ y: y \leq 2\}$—or, more simply, $\mathrm{RR} = \{y \leq 2\}$— 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 $RR = \{y \leq 2\}$ 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

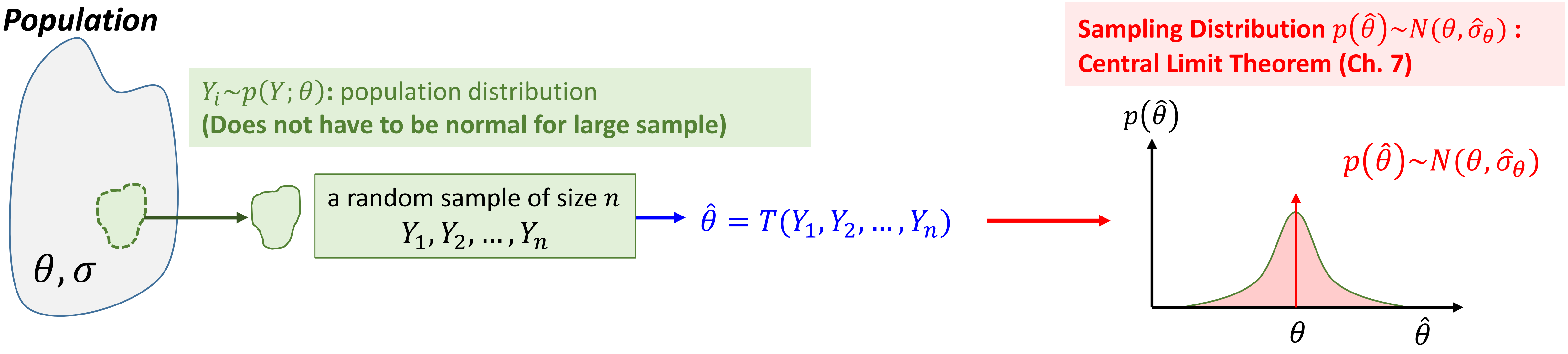
$$\alpha = P(\textit{test statistic is in RR when } H_0 \textit{ is true}) = P(Y \leq 5|p = .5) = \sum_{y=6}^{15} \binom{15}{y} (.5)^{15} = .151.$$

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

$$= .278.$$

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

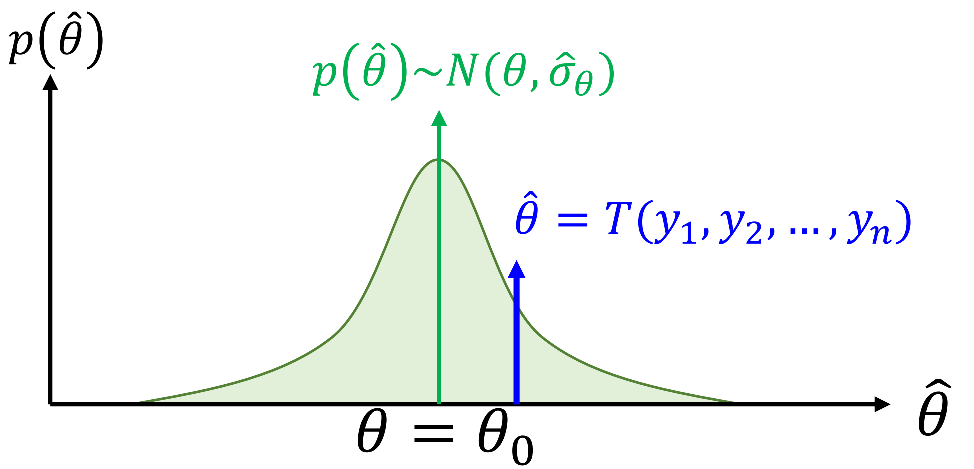
Motivation



- 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 $(p1 - p2)$.

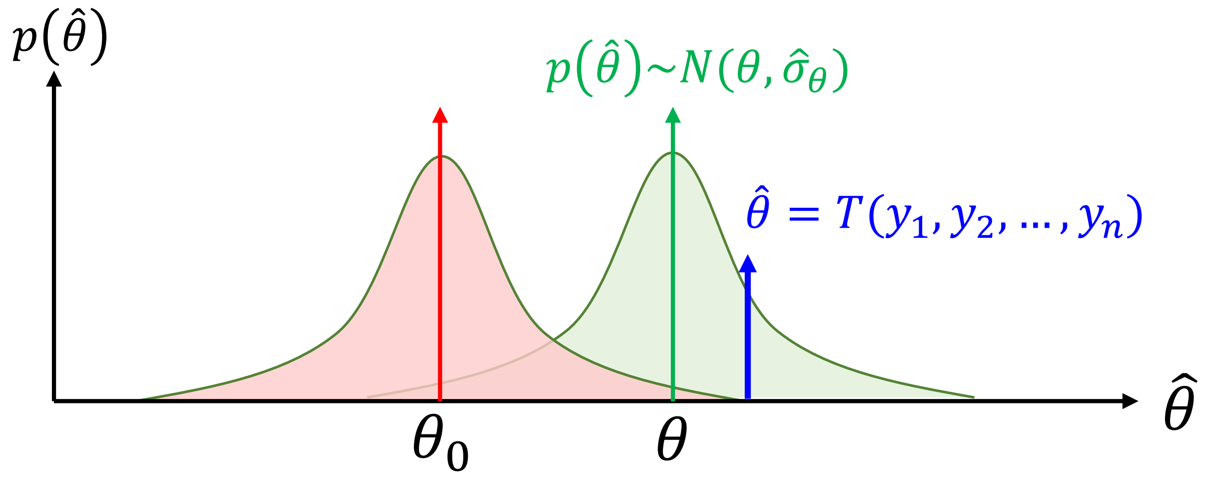
Procedure

- 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 $\Rightarrow$ **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 $\Rightarrow$**Reject** H_0: $\theta = \theta_0$

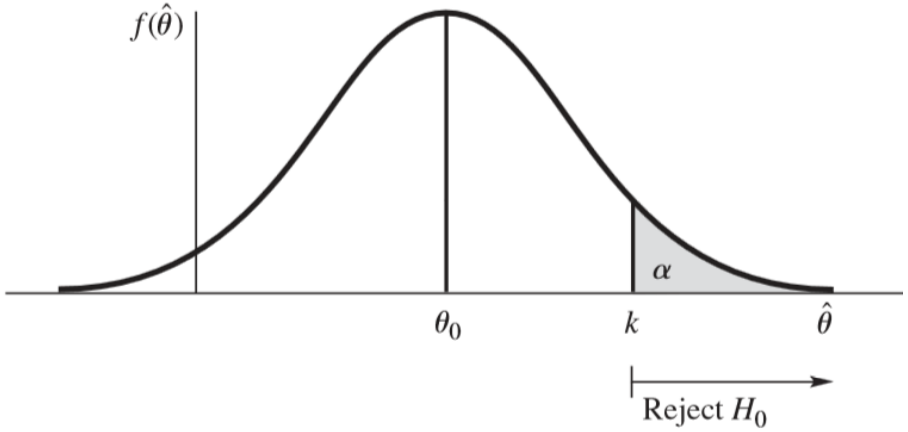
Procedure

- 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

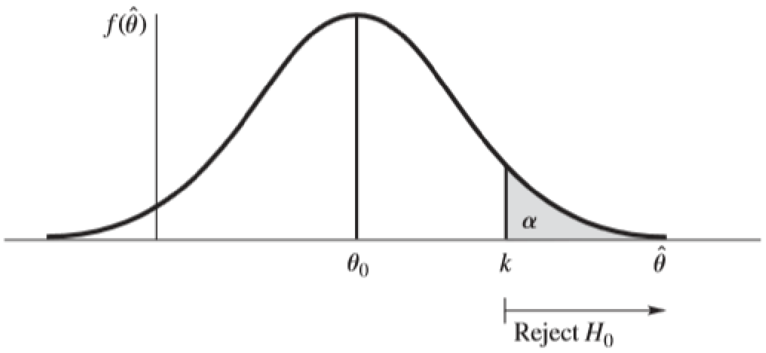
- Large-Sample test
  - ✓ $H_0: \theta = \theta_0$
  - ✓ $H_a: \theta > \theta_0$ (upper tail alternative)
  - ✓ Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
  - ✓ Rejection region:RR= $\{z > z_\alpha\}$: (upper tail rejection region)

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

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

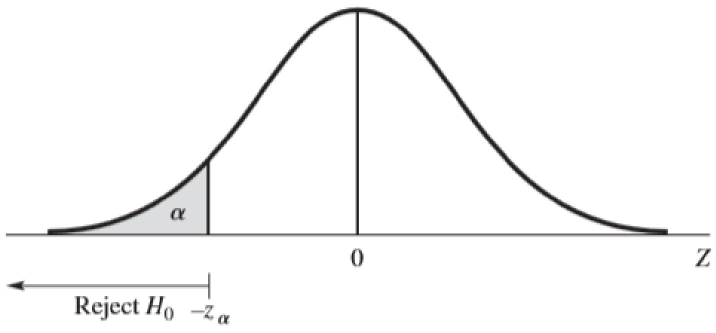
Upper, Lower, Two-tailed Hypothesis Tests

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



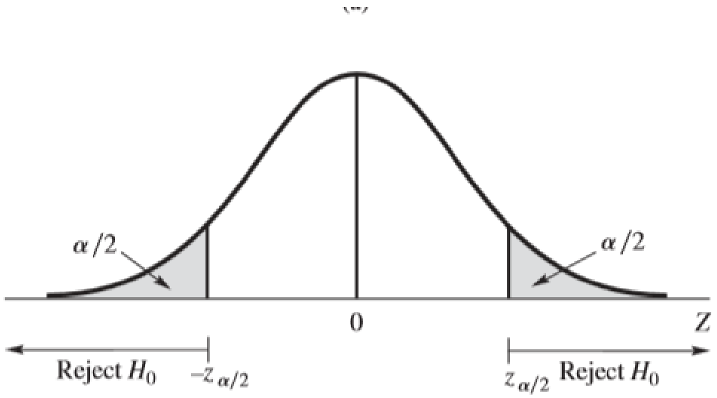
- $H_0: \theta = \theta_0$
- $H_a: \theta > \theta_0$ (upper tail alternative)
- Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- Rejection region:RR$= \{z > z_\alpha\}$ (upper tail rejection region)

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: $Z = \dfrac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- 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: $Z = \dfrac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}$
- 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.

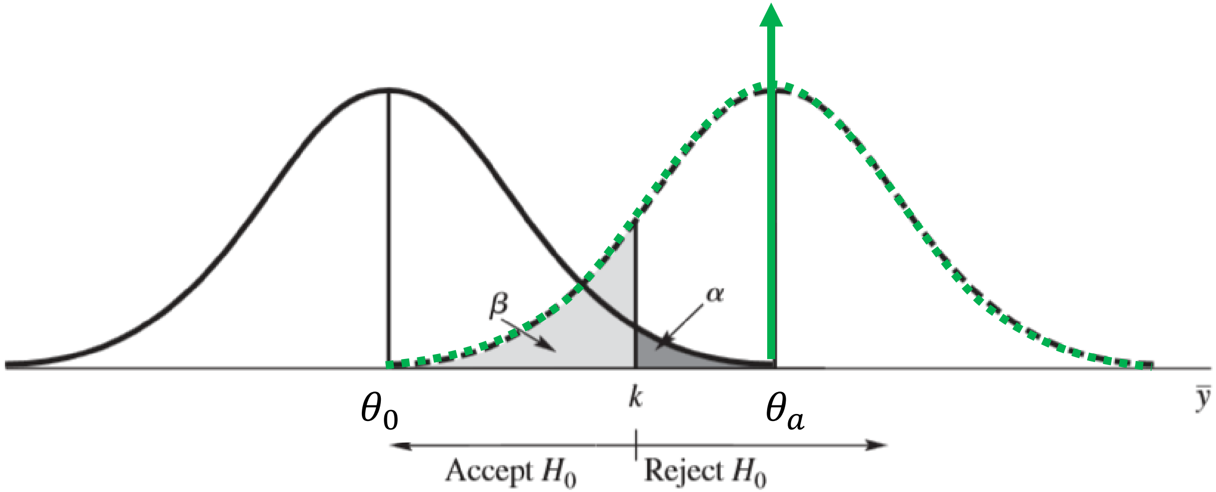
For large sample, sample variances are good estimates of their corresponding population variances, so $z \simeq \frac{3.6 - 3.8}{\sqrt{\frac{.18}{50} + \frac{.14}{50}}} = -2.5$.

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

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



- 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 $\bar{y} > 15.8225$.

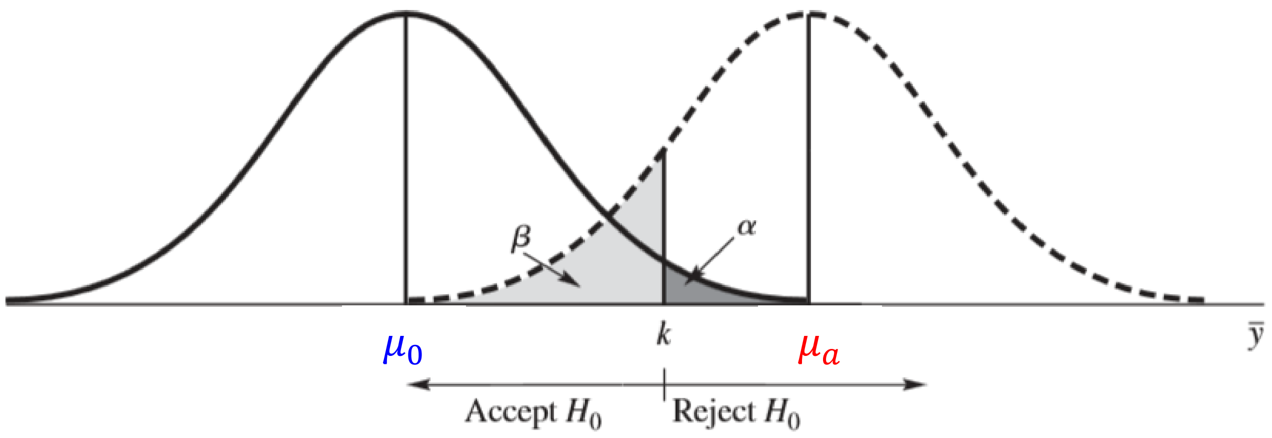
$$\beta = P\left(\frac{\bar{Y}-\mu_a}{\sigma/\sqrt{n}} \leq \frac{15.8225-16}{3/\sqrt{36}}\right) = P(Z \leq -.36) = .3594.$$

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,

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

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



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,

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

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

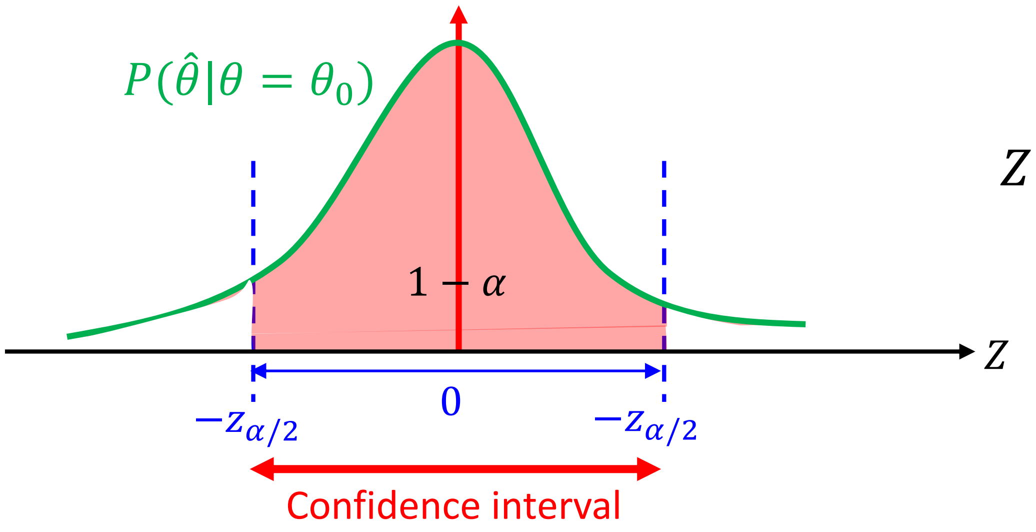
SOLUTION 10.9

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

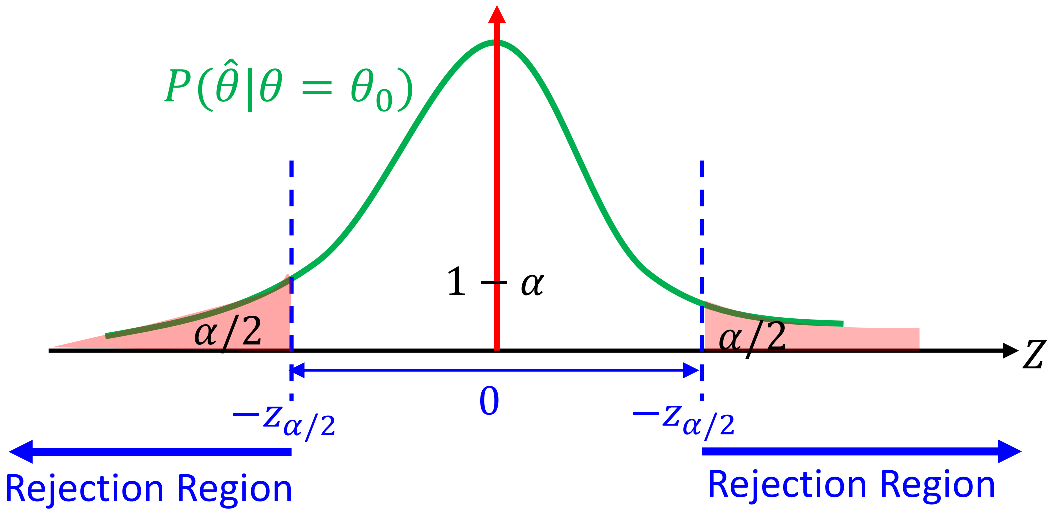
Thus, $n = 98$ observations should be used to meet the requirements.

Motivation

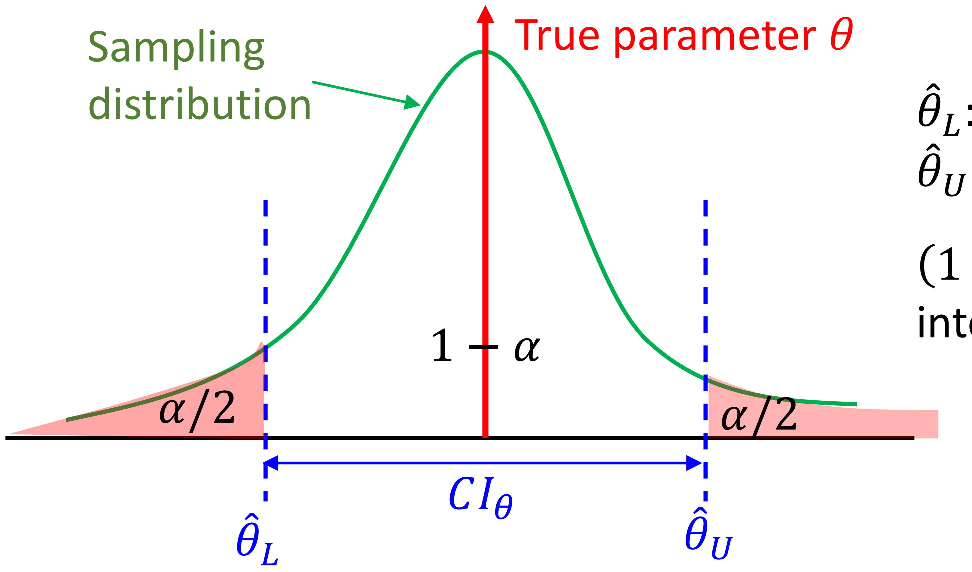
- What is Relationships Between Hypothesis-Testing Procedure?



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



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

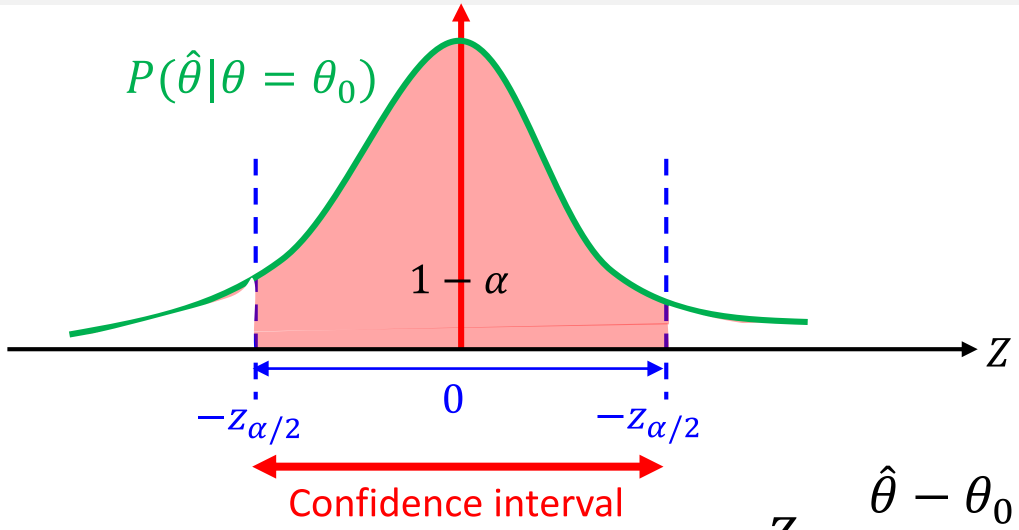
$$P\left(\hat{\theta}_L \le \theta \le \hat{\theta}_L\right) = 1-\alpha$$

- *"There is a $(1-\alpha)$ % probability that when I compute the confidence interval (CI) from a current data sample, the computed CI contains θ*

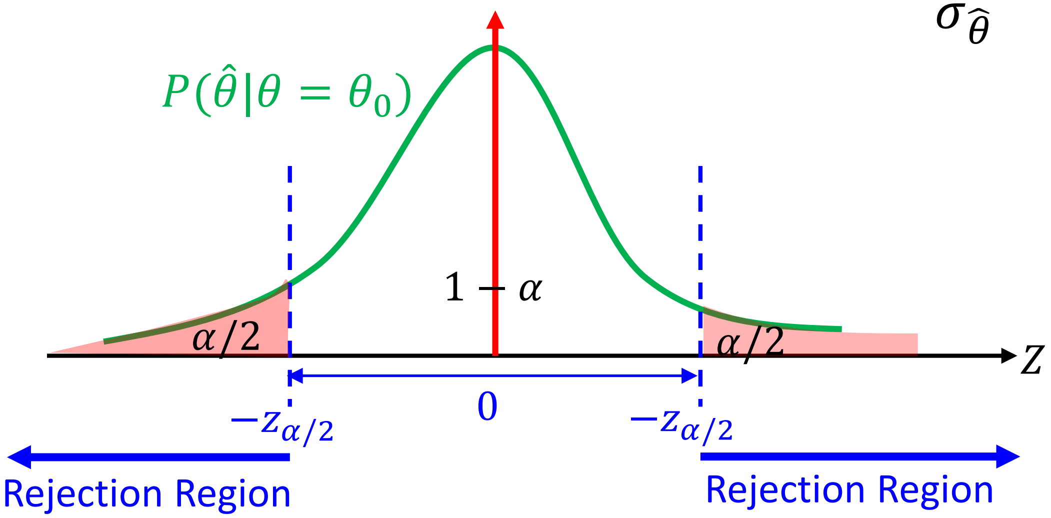
  → From current data set, We can only say that $\theta \in \mathrm{CI}$ or $\theta \notin \mathrm{CI}$

- 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

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

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

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= $\{|z\} > z_{\alpha/2}\}$

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

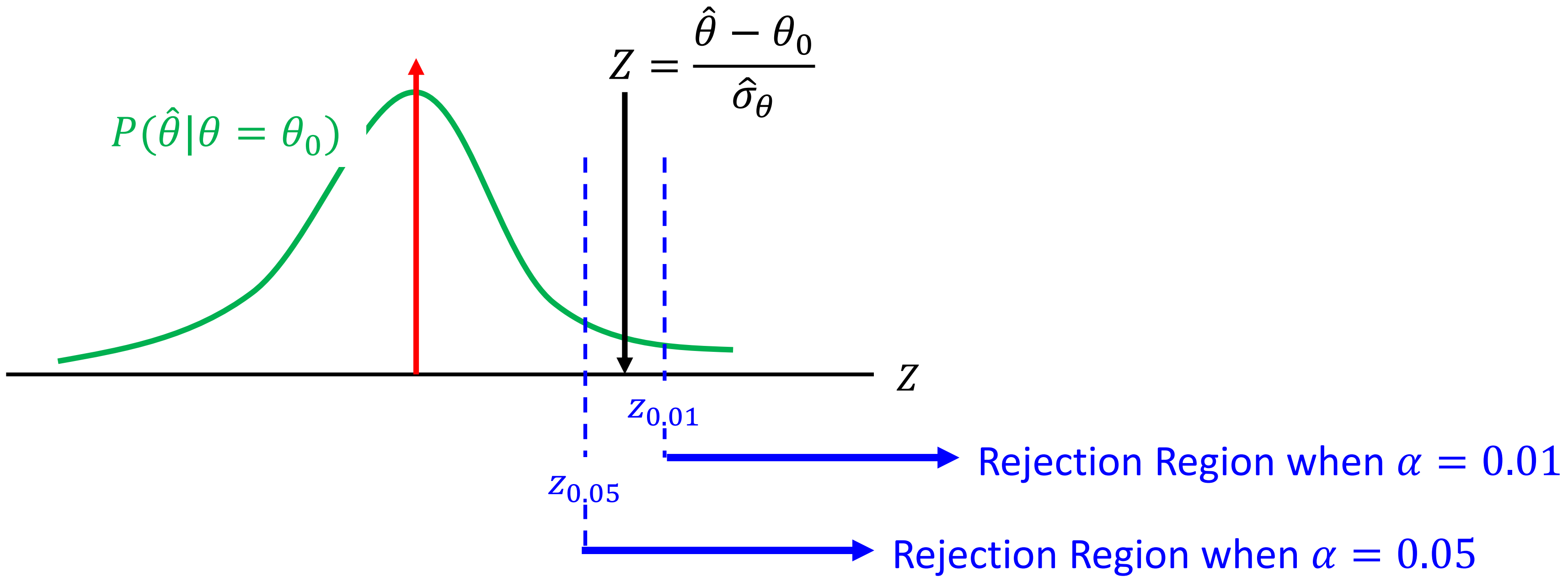
$$P(\theta_0 < \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}} < \theta_0) = \alpha$$

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

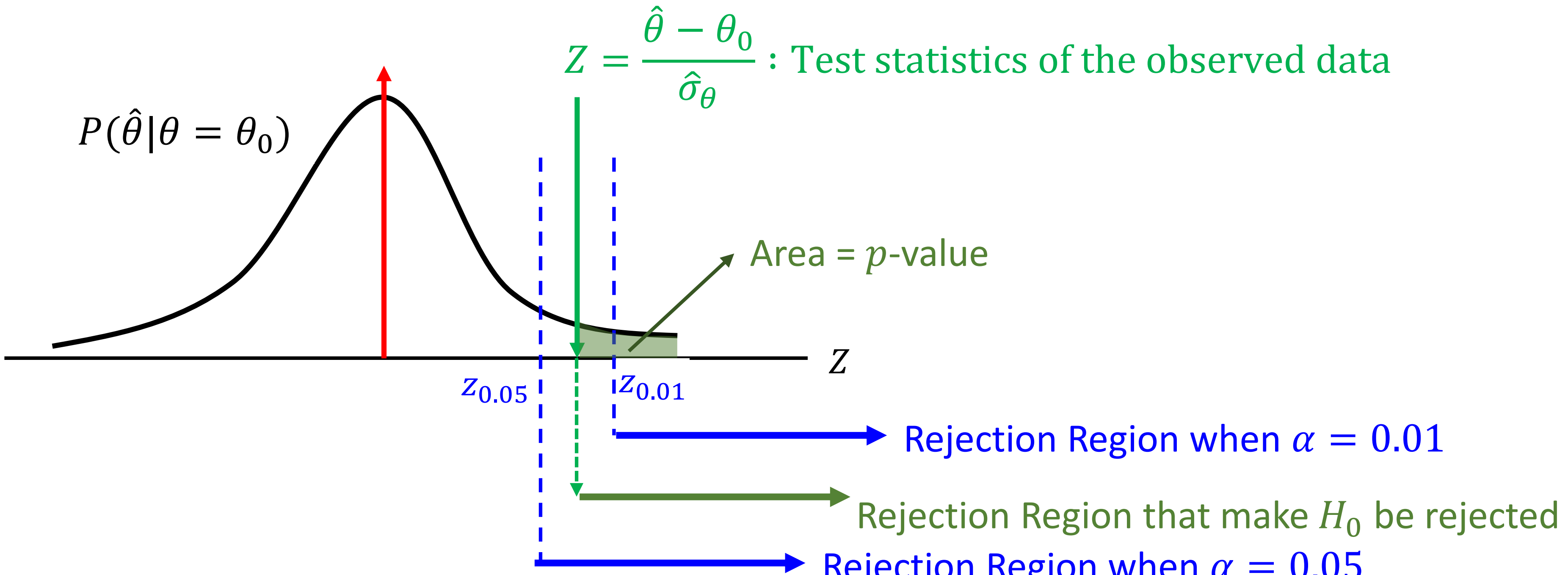


- It is possible, therefore, for two persons to analyze the same data and reach opposite conclusions
  - ✓ one concluding that the null hypothesis should ***be rejected*** at the $\alpha = .05$ significance level
  - ✓ the other deciding that the null hypothesis should ***not be rejected*** with $\alpha = .01$.
- 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

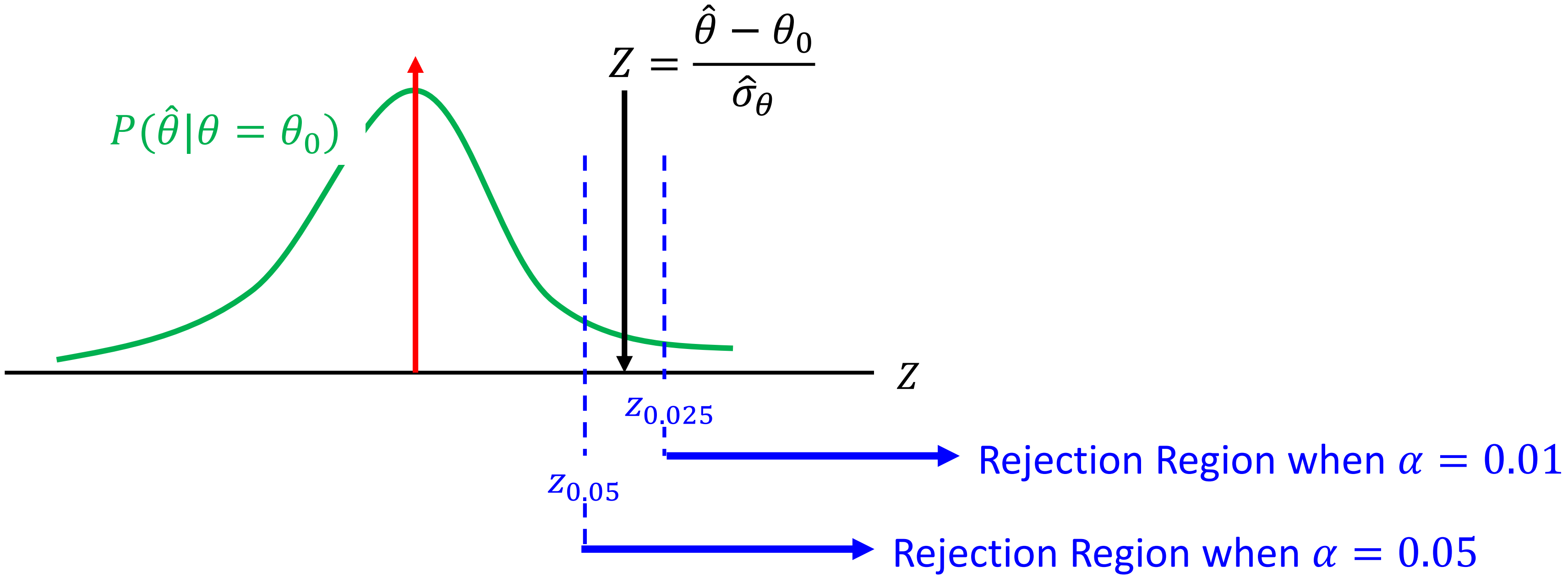
If W is a test statistic, the *p-value*, or *attained significance level*, is **the smallest level** of significance α for which the observed data indicate that the null hypothesis should be rejected.



- 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 $P(Y \leq 3)$, 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 $\alpha \geq .018$ (rejected) , while the null hypothesis could not be rejected for $\alpha \leq .018$.

- Larger α, easier to be rejected

Example

EXAMPLE 10.11

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

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
$s_1^2 = .18$	$s_2^2 = .14$

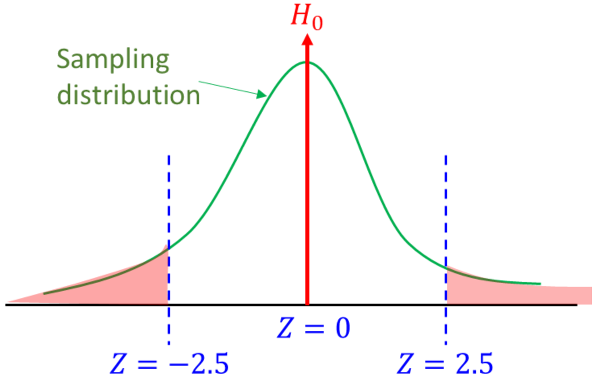
Example

SOLUTION 10.11

Example 10.7 presents a test of $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$.

Test statistic is computed as

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \simeq \frac{3.6 - 3.8}{\sqrt{\frac{.18}{50} + \frac{.14}{50}}} = -2.5,$$



(For large sample, sample variances are good estimates of their corresponding population variances)

The value of the computed test statistic was $z = -2.5$. Because this test is two-tailed, the p-value is

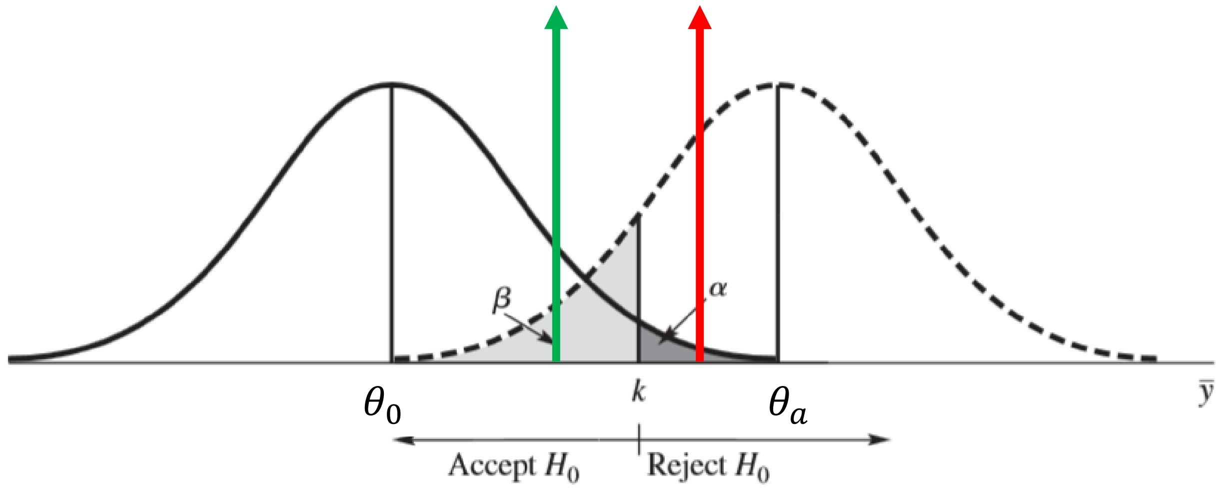
$$P(|Z| \geq 2.5) = P(Z \geq 2.5 \text{ or } Z \leq -2.5) = 2(.0062) = .0124.$$

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

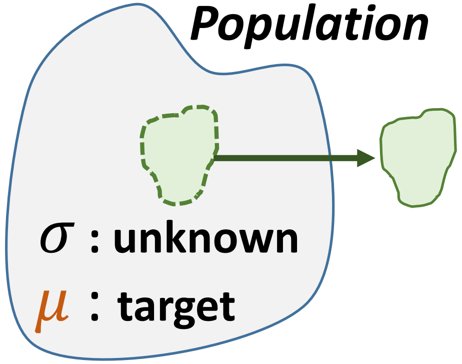
- When the **value of the test statistic** is not in the rejection region, we will "**fail to reject**" rather than "accept" the null hypothesis.
  - If, however, ***Y* does not fall in the rejection region** and we can determine no specific value of θ_a in H_a that is of direct interest, we simply state 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 ***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
$Y_1, Y_2, \ldots, Y_n$

Sample mean

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

Sample variance

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

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

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

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

- 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 $Y_1, Y_2, \dots, Y_n$ 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?

- **Identify null hypothesis.**

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 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 10.13

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.

- **Identify null hypothesis.**

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 $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$. 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

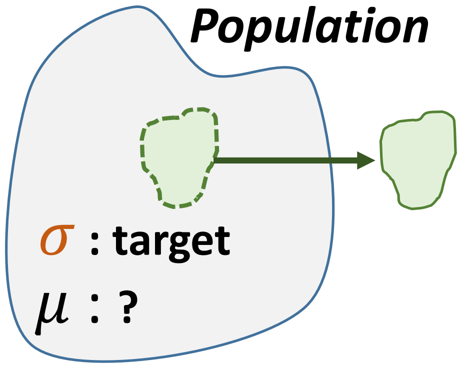
Find the p-value for the statistical test in Example 10.14.

Example

SOLUTION 10.15

The p-value for this test is $P(T > 1.65 \text{ or } T < -1.65)$.
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 $0.10 < p - \text{value} < 0.20$.

Motivation



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

Sample mean

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

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

Sample variance

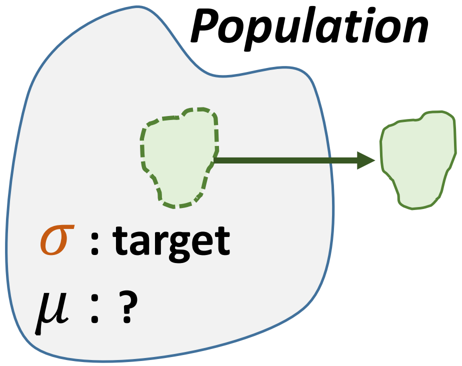
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2, \mathrm{df} = (n-1)$$

- When $Y_i \sim N(\mu, \sigma^2)$
- When n is **small or large**

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

- In this section, we consider the problem of testing $H_0: \sigma^2 = \sigma_0^2$ for some fixed value σ_0^2 versus various alternative hypothesis.

Problem setup



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

Sample mean

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

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

Sample variance

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

- If $H_0: \sigma^2 = \sigma_0^2$ is true,

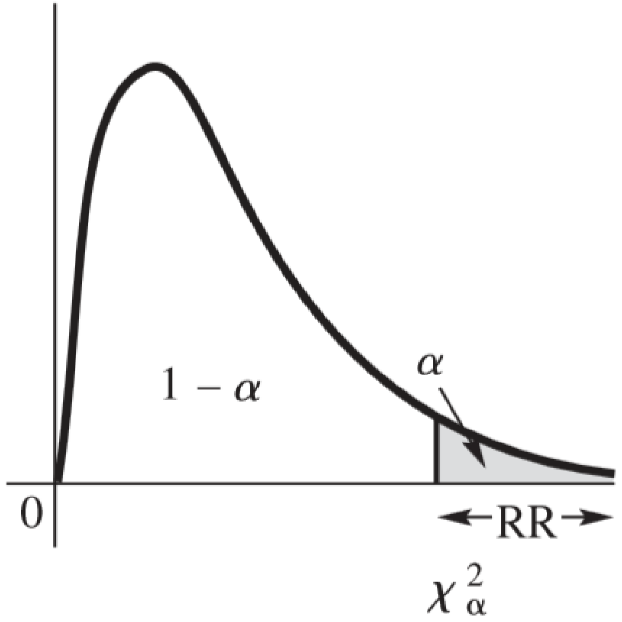
$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2, \mathrm{df} = (n-1)$$

- The four elements of hypothesis testing are:

  - ✓ $H_0: \sigma^2 = \sigma_0^2$
  - ✓ $H_a: \sigma^2 > \sigma_0^2$
  - ✓ Test statistic: $\chi^2 = (n-1)S^2/\sigma_0^2$
  - ✓ Rejection region: RR= $\{\chi^2 > k\}$ for some choice of k

Rejection Region

- The four elements of hypothesis testing are:
  - ✓ $H_0: \sigma^2 = \sigma_0^2$
  - ✓ $H_a: \sigma^2 > \sigma_0^2$
  - ✓ Test statistic: $\chi^2 = (n-1)S^2/\sigma_0^2$
  - ✓ Rejection region: RR$= \{\chi^2 > k\}$ for some choice of k



- If H_a is true and the actual value of σ^2 is larger than σ_0^2
- we would expect S^2 (which estimates the true value of σ^2) to be larger than σ_0^2.
- The larger S^2 is relative to σ_0^2, the stronger is the evidence to support $H_a: \sigma^2 > \sigma_0^2$
- Thus , a rejection region of the form RR$= \{\chi^2 > k\}$ for some constant k is appropriate for testing H_0 versus H_a

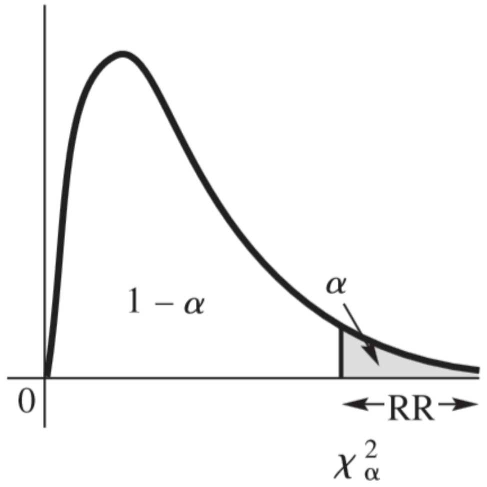
- If we desire a test for which the probability of a ***type I error*** is α, we use the rejection region

$$\text{RR} = \{\chi^2 > \chi^2_\alpha\}$$

where $P(\chi^2 > \chi^2_\alpha) = \alpha$. (Values of χ^2_α can be found in Table 6, Appendix 3.)

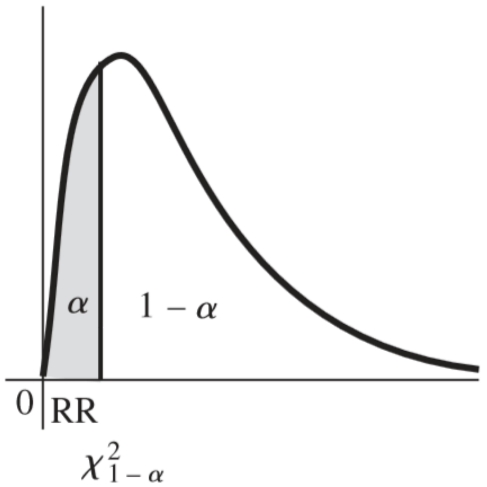
Upper, Lower, Two-tailed Hypothesis Tests

Testing $H_0: \sigma^2 = \sigma_0^2$ against $H_a: \sigma^2 > \sigma_0^2$



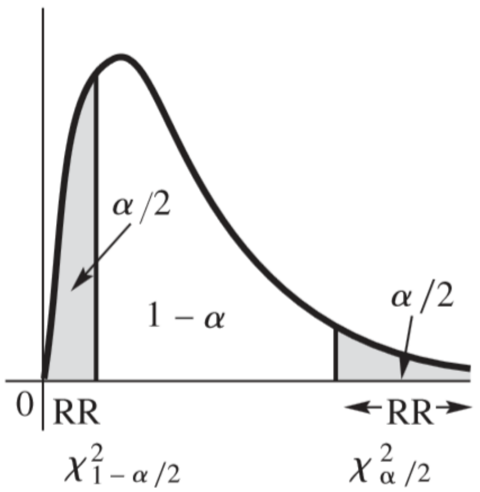
- $H_0: \sigma^2 = \sigma_0^2$
- $H_a: \sigma^2 > \sigma_0^2$ (upper tail alternative)
- Test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$
- Rejection region:RR= $\{\chi^2 > \chi^2_\alpha\}$
  (upper tail rejection region)

Testing $H_0: \sigma^2 = \sigma_0^2$ against $H_a: \sigma^2 < \sigma_0^2$



- $H_0: \sigma^2 = \sigma_0^2$
- $H_a: \sigma^2 < \sigma_0^2$ (lower tail alternative)
- Test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$
- Rejection region:RR= $\{\chi^2 < \chi^2_{1-\alpha}\}$
  (lower tail rejection region)

Testing $H_0: \sigma^2 = \sigma_0^2$ against $H_a: \sigma^2 \neq \sigma_0^2$



- $H_0: \sigma^2 = \sigma_0^2$
- $H_a: \sigma^2 \neq \sigma_0^2$ (two-sided alternative)
- Test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$
- Rejection region:
- RR= $\left\{\chi^2 > \chi^2_{\alpha/2} \text{ or } \chi^2 < \chi^2_{1-\alpha/2}\right\}$:
  (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 10.16

A company produces machined engine parts that are supposed to have a diameter variance no larger than .0002 (diameters measured in inches). A random sample of ten parts gave a sample variance of .0003. Test, at the 5% level, $H_0\colon \sigma^2 = .0002$ against $H_a\colon \sigma^2 > .0002$.

Example

SOLUTION 10.16

The appropriate test statistic is $\chi^2 = (n-1)S^2/\sigma_0^2$ if it is reasonable to assume that the measured diameters are normally distributed.

Because we have posed an upper-tail test, the rejection region is $\chi^2 > \chi^2_{0.05} = 16.919$ (based on 9 df).

The observed value of test statistic is

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{(9)(.0003)}{.0002} = 13.5.$$

Thus, the null hypothesis is not rejected.

Example

EXAMPLE 10.17

Determine the p-value associated with the statistical test of Example 10.16.

Example

SOLUTION 10.17

The p-value is $P(\chi^2 > 13.5)$ where χ^2 is based on 9 df. By examining the row corresponding to 9 df in Table 6, Appendix 3, we find that $\chi^2_{.1} = 14.6837$. Thus the p-value > 0.1.

Example

EXAMPLE 10.18

An experimenter was convinced that the variability in his measuring equipment results in a standard deviation of 2. Sixteen measurements yielded $s^2 = 6.1$. Do the data disagree with his claim? Determine the *p*-value for the test. What would you conclude if you chose $\alpha = .05$?

Example

SOLUTION 10.18

We require a test of $H_0: \sigma^2 = 4$ versus $H_a: \sigma^2 \neq 4$, a two-tailed test.

The value of the test statistic is $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = 15(6.1)/4 = 22.875$.

Referring to Table 6, Appendix 3, we see tat for 15 df, $\chi^2_{.05} = 24.9958$ and $\chi^2_{.10} = 22.3072$.

Thus, the portion of the p-value that falls in the upper tail is between .05 and .10.

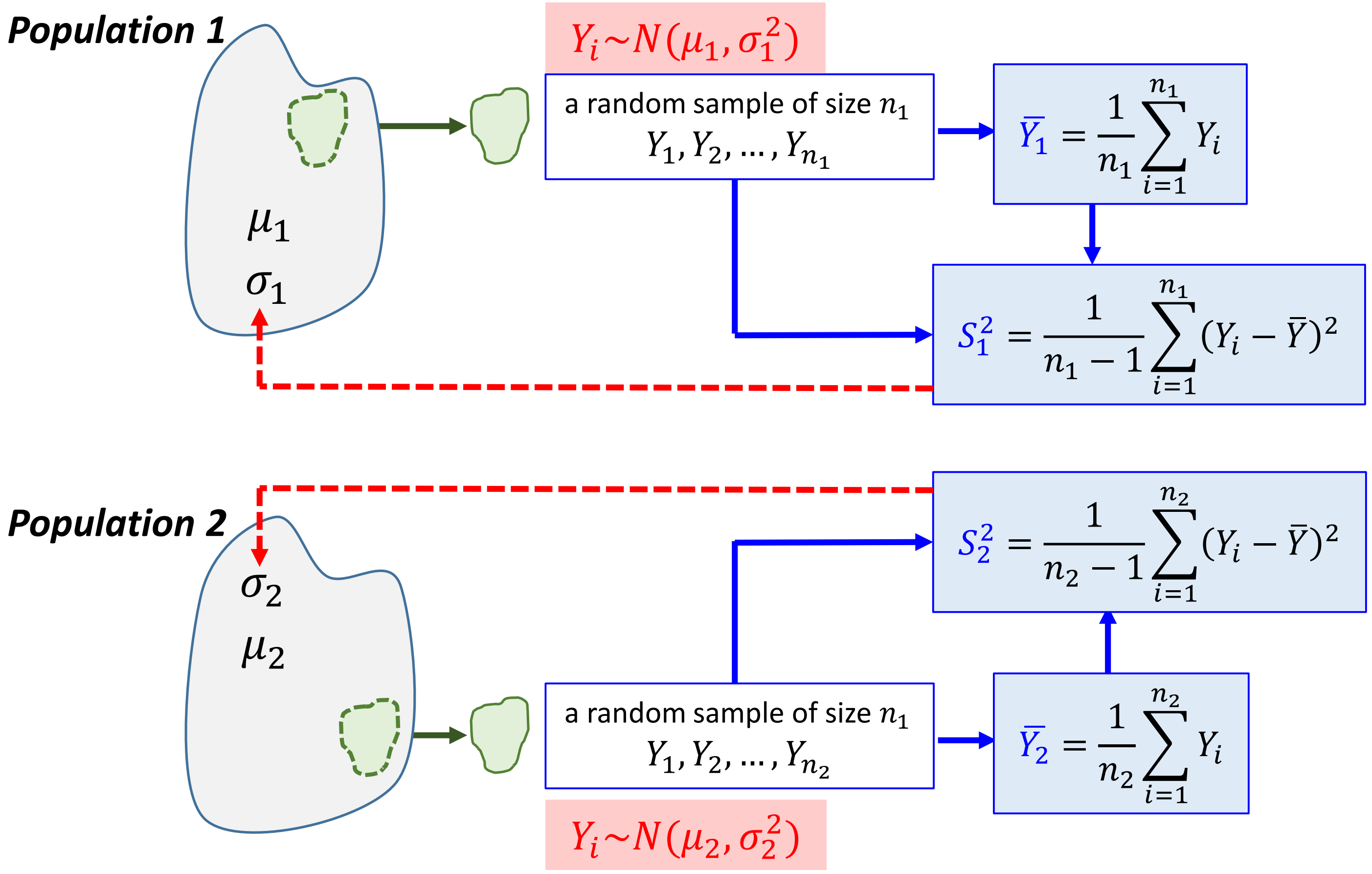
Because we need to account for a corresponding equal area in the lower tail, it follows that $.1 < p-$ value $< .2$. It is clear that the chosen value of $\alpha = .05$ is smaller than the p-value, therefore we cannot reject the null hypothesis.

Comparing Variances

- Sometimes we wish to compare the variances of two normal distributions, particularly by testing to determine whether they are equal.

- These problems are encountered in comparing
  - ✓ the precision of two measuring instruments,
  - ✓ the variation in quality characteristics of a manufactured product, or
  - ✓ the variation in scores for two testing procedures.

Comparing Variances



$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right) \sim F$$

with $(n_1 - 1)$ df for numerator
$(n_2 - 1)$ df for denominator

- Thus, it seems intuitive that the ratio S_1^2/S_2^2 could be used to make inferences about the relative magnitudes of σ_1^2 and σ_2^2.
- Suppose that we want to test the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ against $H_a: \sigma_1^2 > \sigma_2^2$

Comparing Variances

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ against } H_a: \sigma_1^2 > \sigma_2^2$$

- Because the sample variances S_1^2 and S_2^2 estimate the respective population variances, we reject H_0 in favor of H_a if S_1^2 is much larger than S_2^2. That is, we use a rejection region RR of the form

$$\text{RR} = \left\{ \frac{S_1^2}{S_2^2} > k \right\}$$

  - ✓ Where k is chosen so that the probability of a ***type I error*** is α.
  - ✓ The appropriate value of k depends on the probability distribution of the statistic S_1^2/S_2^2

- Chapter 7 has shown that

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right) \sim F$$

with $(n_1 - 1)$ df for numerator
$(n_2 - 1)$ df for denominator

- Under the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$, $S_1^2/S_2^2 \sim F$

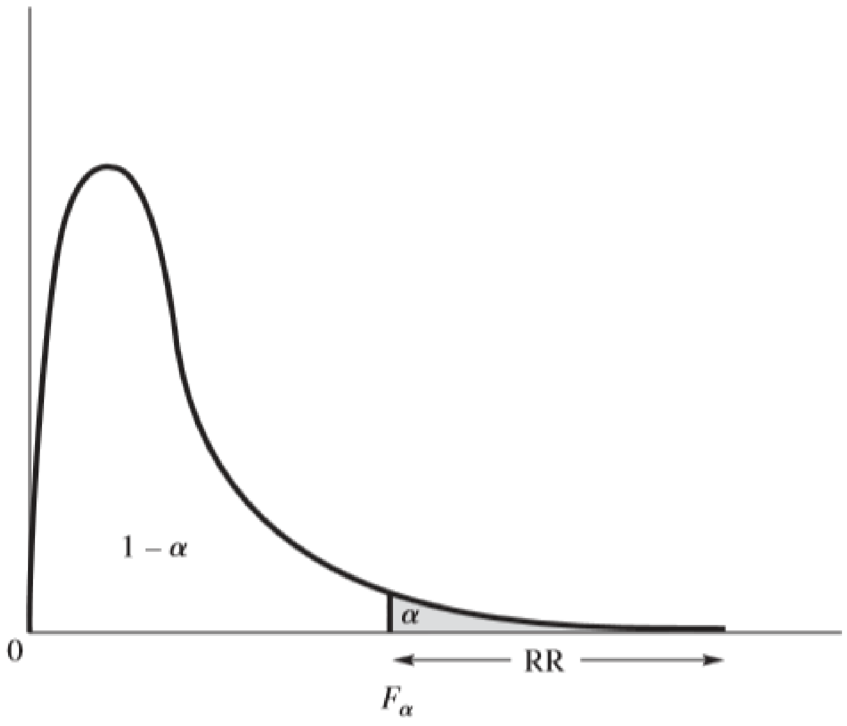
$$\text{RR} = \left\{ \frac{S_1^2}{S_2^2} > k \right\} = \{F > k\} = \{F > F_\alpha\}$$

where $k = F_\alpha$ is the value of F distribution $v_1 = (n_1 - 1)$ and $v_2 = (n_2 - 1)$ such that $P(F > F_\alpha) = \alpha$

Summary and Remarks

Test of the Hypothesis $\sigma_1^2 = \sigma_2^2$

Assumptions: Independent samples from normal populations.
$H_0: \sigma_1^2 = \sigma_2^2$.
$H_a: \sigma_1^2 > \sigma_2^2$.
Test statistic: $F = \dfrac{S_1^2}{S_2^2}$.
Rejection region: $F > F_\alpha$, where F_α is chosen so that $P(F > F_\alpha) = \alpha$ when F has $\nu_1 = n_1 - 1$ numerator degrees of freedom and $\nu_2 = n_2 - 1$ denominator degrees of freedom. (See Table 7, Appendix 3.)



- Both the χ^2 tests and the F tests presented in this section are very sensitive to departures from the assumption of **normality of the underlying population(s).**
- Thus, unlike the tests of Section 10.8, **these tests are not robust** if the normality assumption is violated.

Example

EXAMPLE 10.19

Suppose that we wish to compare the variation in diameters of parts produced by the company in Example 10.16 with the variation in diameters of parts produced by a competitor. Recall that the sample variance for our company, based on $n = 10$ diameters, was $s_1^2 = .0003$. In contrast, the sample variance of the diameter measurements for 20 of the competitor's parts was $s_2^2 = .0001$. Do the data provide sufficient information to indicate a smaller variation in diameters for the competitor? Test with $\alpha = .05$.

Example

SOLUTION 10.19

We are testing $H_0\colon \sigma_1^2 = \sigma_2^2$ against $H_a\colon \sigma_1^2 > \sigma_2^2$.

The test statistic $F = S_1^2/S_2^2$ is based on $\nu_1 = 9$ numerator and $\nu_2 = 19$ denominator df, and we reject H_0 for values of F larger than $F_{.05} = 2.42$. (See Table 7, Appendix 3.)

Because the observed value of the test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{.0003}{.0001} = 3,$$

we see that $F > F_{.05}$ and reject the null hypothesis.

Example

EXAMPLE 10.20

Give bounds for the p-value associated with the data of Example 10.19.

Example

SOLUTION 10.20

The calculated F value for this upper-tail test is $F = 3$. Because this value is based on $\nu_1 = 9$ and $\nu_2 = 19$ numerator and denominator df, Table 7, Appendix 3, can be used to determine that $F_{.025} = 2.88$ whereas $F_{.01} = 3.52$. Thus, $.01 < p - \text{value} < .025$.

Example

EXAMPLE 10.21

An experiment to explore the pain thresholds to electrical shocks for males and females resulted in the data summary given in Table 10.4. Do the data provide sufficient evidence to indicate a significant difference in the variability of pain thresholds for men and women? Use $\alpha = .10$. What can be said about the p-value?

Table 10.4 Data for Example 10.21

	Males	Females
n	14	10
$\overline{y}$	16.2	14.9
s^2	12.7	26.4

Example

SOLUTION 10.21

Assume that the pain thresholds for men and women are approximately normally distributed. We desire to test $H_0: \sigma_M^2 = \sigma_F^2$ versus $H_a: \sigma_M^2 \neq \sigma_F^2$.

The larger S^2 is 26.4, and the associated sample size is 10. The smaller S^2 is 12.7 and the associated sample size is 14. Therefore,

$$F = \frac{26.4}{12.7} = 2.079.$$

Because $F_{.05} = 2.71$ with $\nu_1 = 10 - 1 = 9$, and $\nu_2 = 14 - 1 = 13$, and 2.079 is not larger than the critical value, we cannot reject the null hypothesis.

Referring to Table 7, Appendix 3, with $\nu_1 = 9, \nu_2 = 13$ numerator and denominator df, we find $F_{.10} = 2.16$. Thus, $p - \text{value} > 2(.10) = .20$.

Motivation

- In the remaining sections of this chapter, we move from practical examples of statistical tests to a ***theoretical discussion*** of their properties.
- We have suggested specific tests for a number of practical hypothesis testing situations, but you may wonder why we chose those particular tests.
  - How did we decide on the ***test statistics*** that were presented?
  - How did we know that we had selected the ***best rejection regions***?
- The goodness of a test is measured by α and β, the probabilities of ***type I*** and ***type II*** errors, respectively.
- Typically, the value of α is chosen in advance and determines the location of the rejection region

Definition

DEFINITION 10.3

Suppose that W is the test statistic and RR is the rejection region for a test of a hypothesis involving the value of a parameter θ. Then the *power* of the test, denoted by $\text{power}(\theta)$, is the probability that the test will lead to rejection of H_0 when the actual parameter value is θ. That is,

$$\text{power}(\theta) = P(W \text{ in RR when the parameter value } is\ \theta).$$

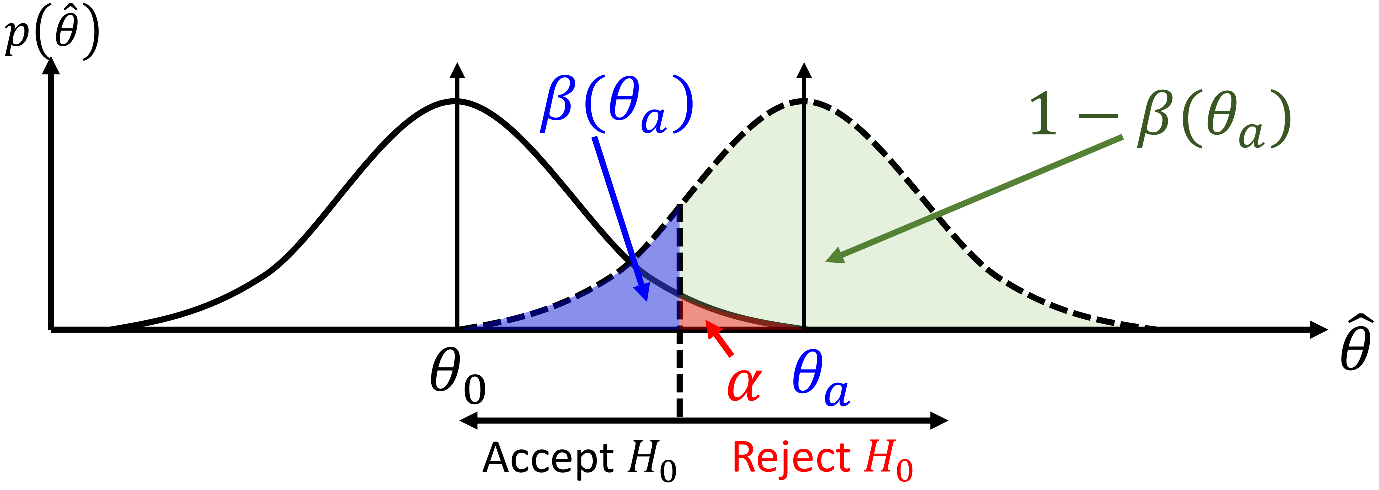
- Suppose that we want to test the null hypothesis $H_0\!: \theta = \theta_0$ and that θ_a is a particular value for θ chosen from H_a.
- The power of the test at $\theta = \theta_0$, power(θ_a), is equal to the probability of rejecting H_0 when H_0 is true.
- That is, power(θ_0), = α, the probability of a ***type I error***.

Relationship Between Power and β

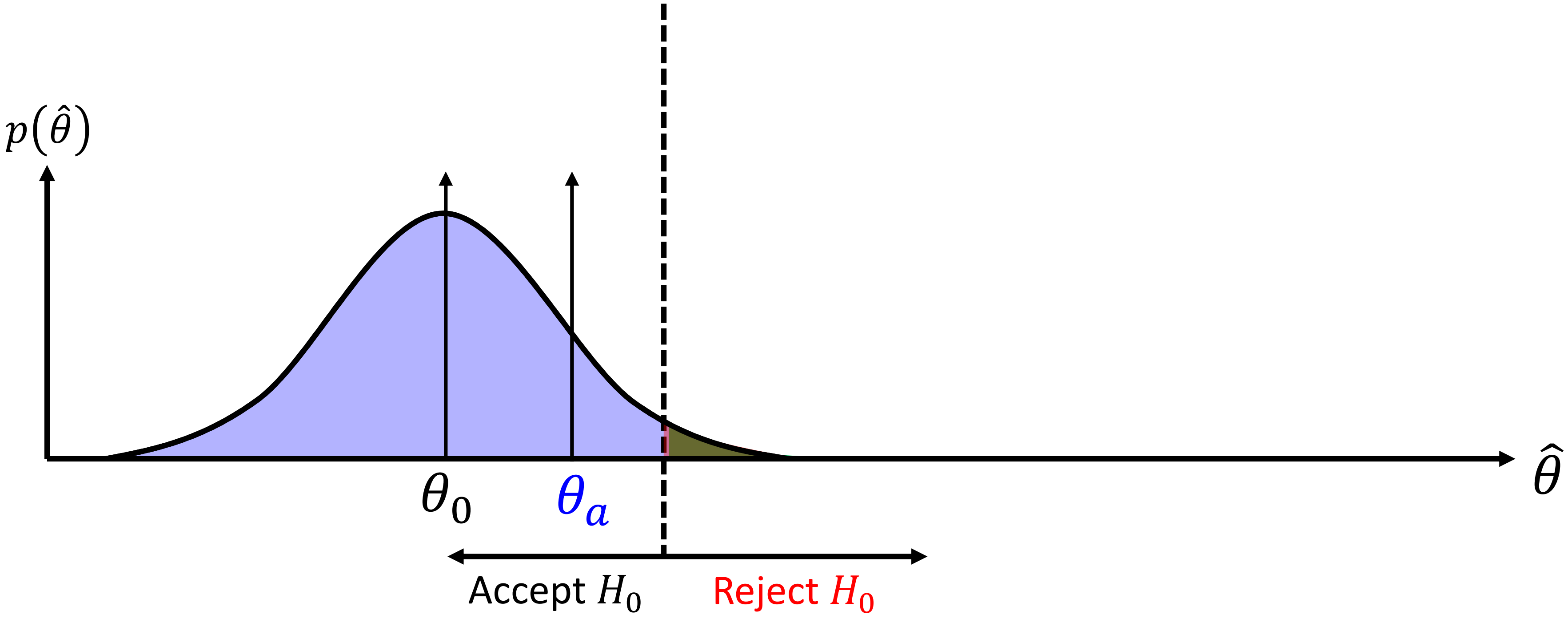
- For any value of θ from H_a, the power of a test measures the ability to detect that the null hypothesis is false.
- That is, for $\theta = \theta_a$,

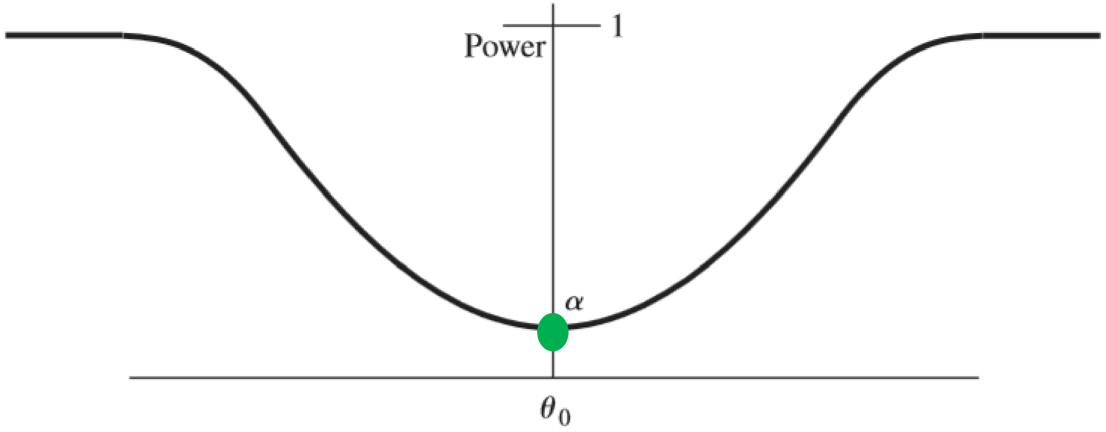
$$\text{power}(\theta_a) = P(\textbf{reject } H_0 \text{ when } \theta = \theta_a) = 1 - P(\textbf{accept } H_0 \text{ when } \theta = \theta_a) = 1 - \beta(\theta_a)$$

	$H_0: \theta = \theta_0$ is True	H_0 is False ($H_a: \theta = \theta_a$ true)
Reject H_0	***type I error*** $P(\textit{type I error}) = \alpha$	**Correct Decision** $1 - \beta(\theta_a)$
Accept H_0	**Correct Decision** $1 - \alpha$	***type I error*** $P(\textit{type II error}) = \beta(\theta_a)$

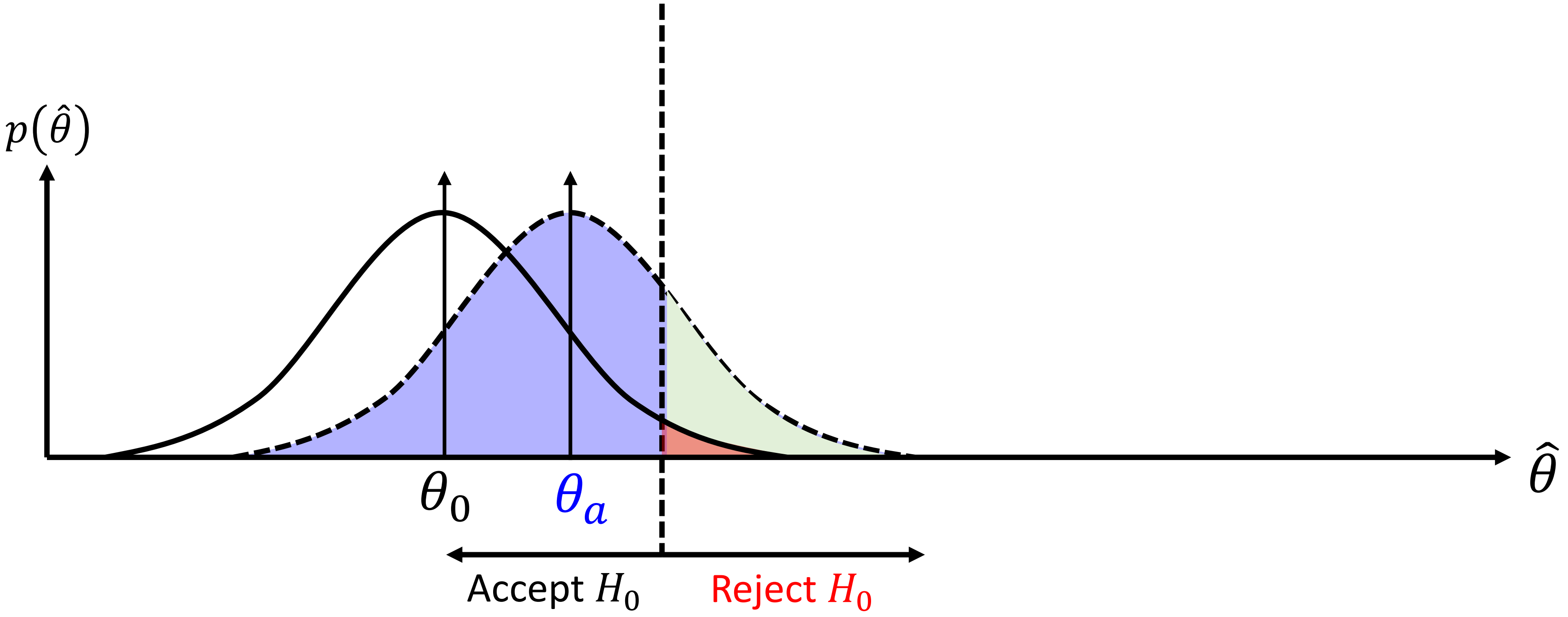


Power curve

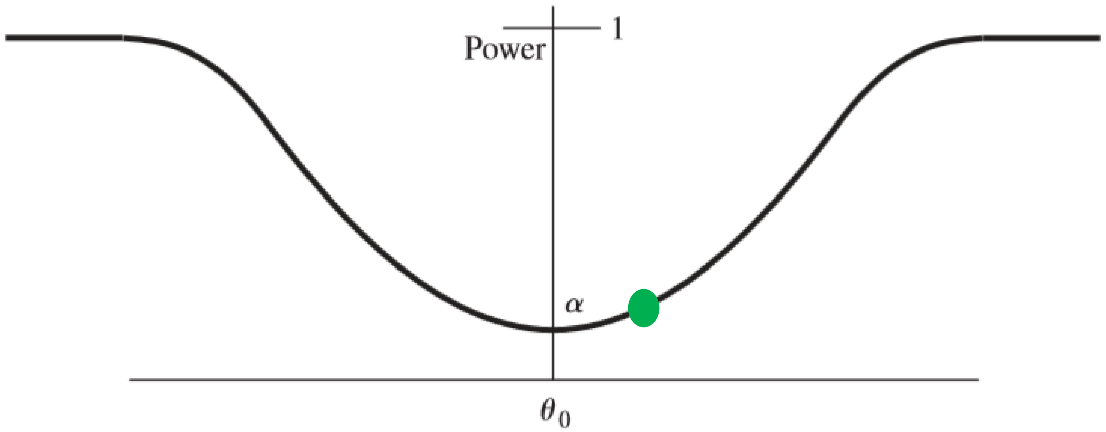

$p(\hat{\theta})$
θ_0
θ_a
$\hat{\theta}$
Accept H_0
Reject H_0


Power
1
α
θ_0

Power curve

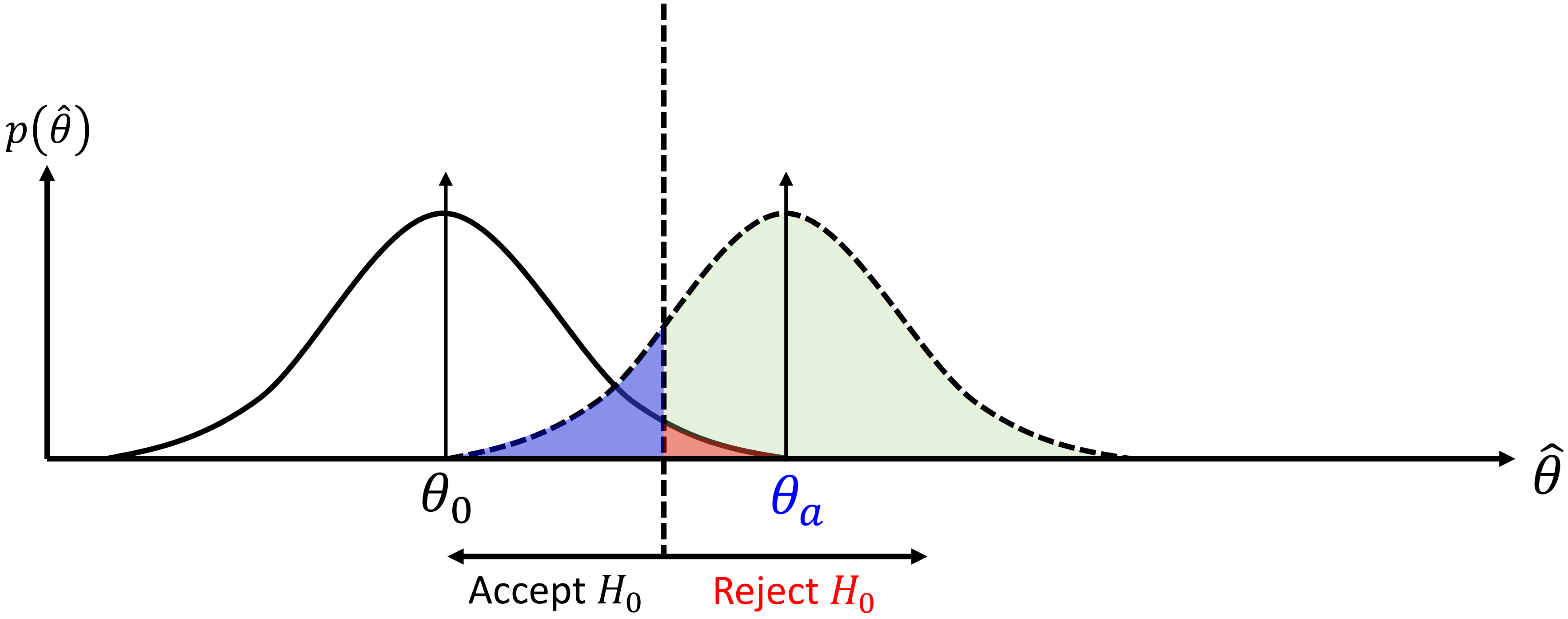


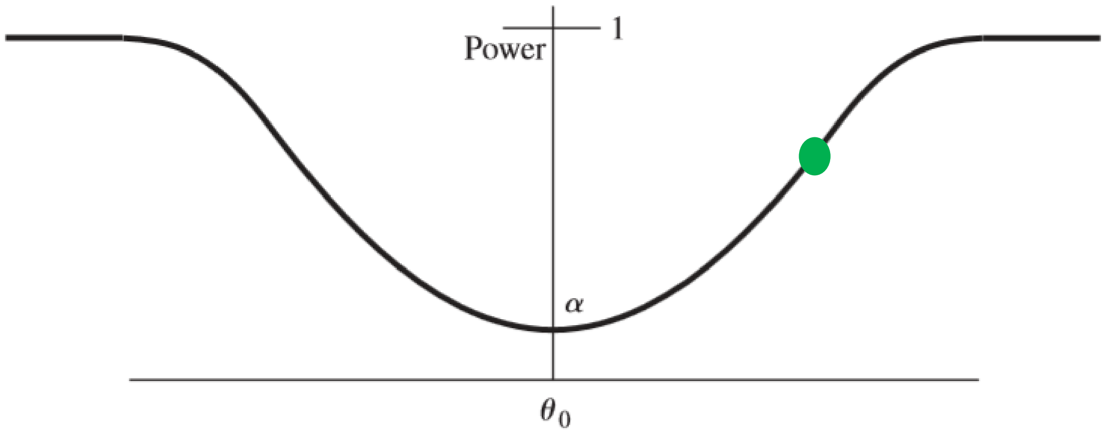
$p(\hat{\theta})$
$\hat{\theta}$
θ_0
θ_a
Accept H_0
Reject H_0



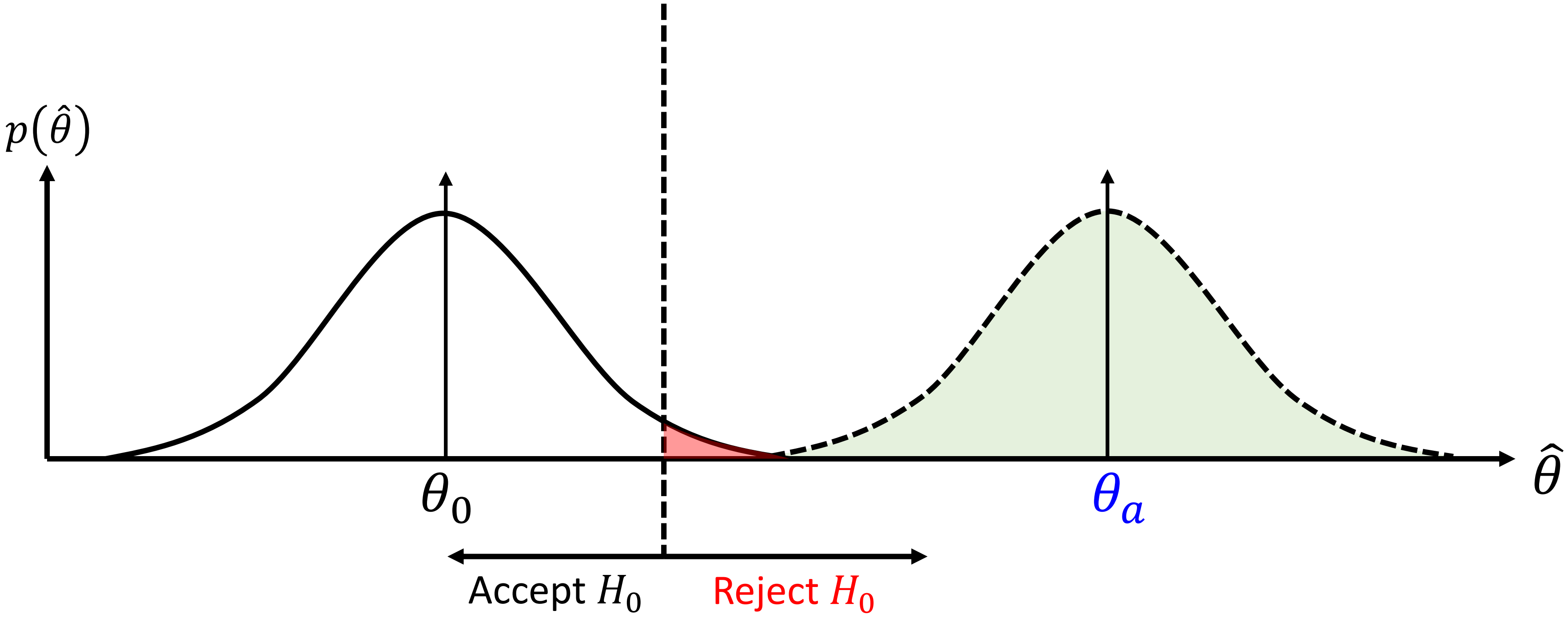
Power
1
α
θ_0

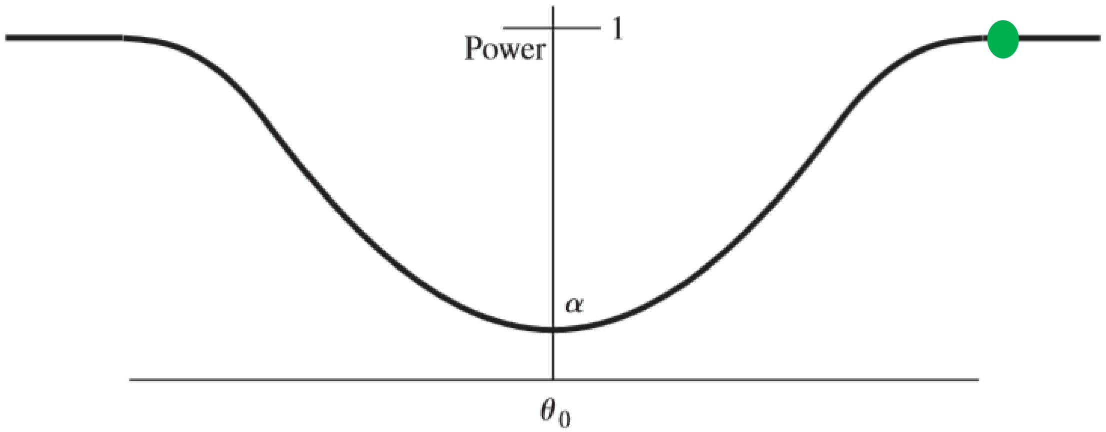
Power curve


$p(\hat{\theta})$
$\hat{\theta}$
θ_0
θ_a
Accept H_0
Reject H_0

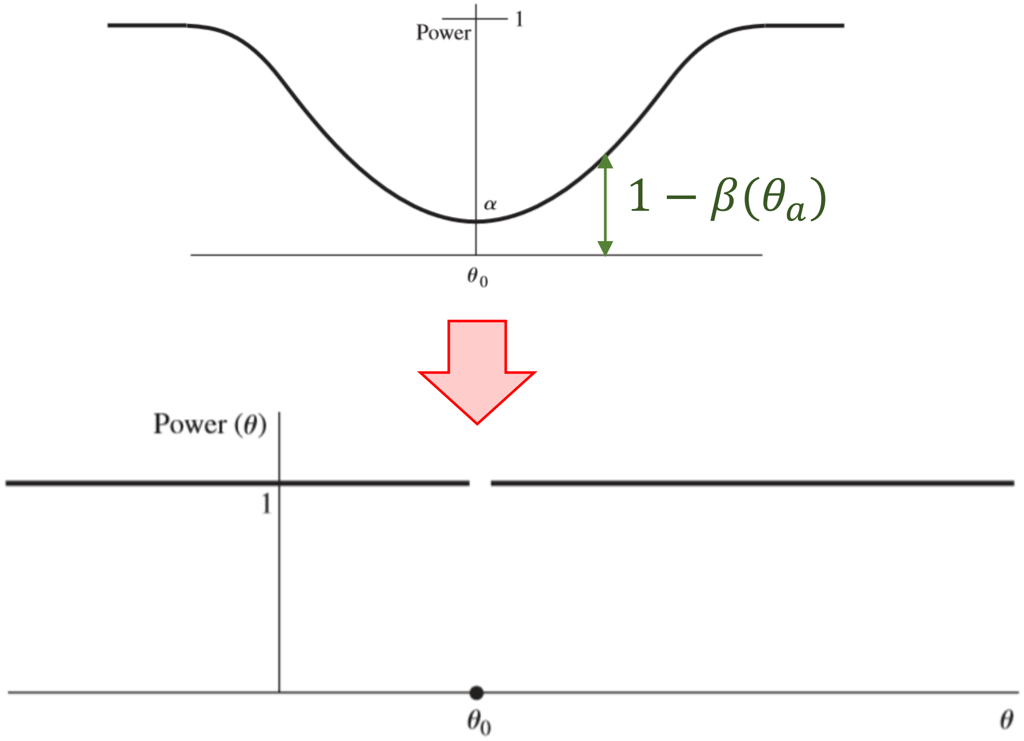

Power
1
α
θ_0

Power curve


$p(\hat{\theta})$
θ_0
θ_a
$\hat{\theta}$
Accept H_0
Reject H_0


Power
1
α
θ_0

Ideal Power Curve



- For a fixed sample size n, we adopt the procedure of selecting a (small) value for α and finding a rejection region RR to minimize $\beta(\theta_a)$ at each θ_a in H_a. Equivalently, we choose RR to maximize $\text{power}(\theta)$ for θ in H_a.
- From among all tests with a significance level of α, we seek the test whose power function comes closest to the ideal power function. If such a test exists. How do we find such a testing procedure?

Simple Hypothesis

DEFINITION 10.4

If a random sample is taken from a distribution with parameter θ, a hypothesis is said to be a ***simple hypothesis*** if that hypothesis *uniquely specifies* the distribution of the population from which the sample is taken. Any hypothesis that is not a simple hypothesis is called a ***composite hypothesis***.

- Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from an exponential distribution with parameter λ

$$f(y) = (1/\lambda)e^{-y/\lambda}, y$$

  - ✓ $H: \lambda = 2$ is simple hypothesis because it uniquely specifies the distribution from which the sample is taken
  - ✓ $H^*: \lambda > 2$ is composite hypothesis because under H^* the density function $f(y)$ is not uniquely determined.

most powerful α level test

- Suppose that we would like to test a simple null hypothesis $H_0: \theta = \theta_0$ versus a simple alternative hypothesis $H_a: \theta = \theta_a$.
- Because we are concerned only with two particular values of θ (θ_0 and θ_a), we would like to choose a rejection region RR so that $\alpha = \text{power}(\theta_0)$ is a fixed value and $\text{power}(\theta_a)$ is as large as possible.
- That is, we seek a most powerful α level test. The following theorem provides the methodology for deriving the most powerful test for testing simple H_0 versus simple H_a

The Neyman-Pearson Lemma

THEOREM 10.1 (The Neyman-Pearson Lemma)

Suppose that we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative hypothesis $H_a : \theta = \theta_a$, based on a random sample $Y_1, Y_2, \ldots, Y_n$ from a distribution with parameter θ. Let $L(\theta)$ denote the likelihood of the sample when the value of the parameter is θ. Then, for a given α, the test that maximizes the power at θ_a has a rejection region, RR, determined by

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

The value of k is chosen so that the test has the desired value for α. Such a test is a most powerful α-level test for H_0 versus H_a.

Example

EXAMPLE 10.22

Suppose that Y represents a single observation from a population with probability density function given by

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1 \\ 0, & elsewhere. \end{cases}$$

Find the most powerful test with significance level $\alpha = .05$ to test $H_0\!: \theta = 2$ versus $H_a\!: \theta = 1$.

Example

SOLUTION 10.22

Because both of the hypotheses are simple, Theorem 10.1 can be applied to derive the required test.

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{f(y|\theta_0)}{f(y|\theta_a)} = 2y, \text{ for } 0 < y < 1.$$

And the form of the rejection region for the most powerful test is $2y < k$ or equivalently, $\mathrm{RR} = \{y < k^*\}$. Because $\alpha = .05$,

$$P(Y \text{ in RR when } \theta = 2) = P(Y < k^* | \theta = 2) = \int_0^{k^*} 2y dy = (k^*)^2.$$

Therefore, $k^* = \sqrt{.05} (\because y > 0)$ and $RR = \{y < \sqrt{.05} = .2236\}$

- Notice that the forms of the test statistic and of the rejection region depend on both H_0 and H_a.
- If the alternative is changed to $H_a : \theta = 4$, the most powerful test is based on Y^2, and we reject H0 in favor of H_a if $Y^2 > k'$, for some constant k'.
- Also notice that the Neyman–Pearson lemma gives the form of the rejection region; the actual rejection region depends on the specified value for α.

Uniformly most powerful test

- If we desire to test $H_0: \theta = \theta_0$ (**simple**) versus$H_a: \theta > \theta_0$ (**composite**), no general theorem comparable to Theorem 10.1 is applicable if either hypothesis is composite.

- However, Theorem 10.1 can be applied to obtain a most powerful test for $H_0: \theta = \theta_0$versus $H_a: \theta = \theta_a$ for any single value θ_a, where $\theta_a > \theta_0$.

- In many situations, the actual rejection region for the most powerful test depends only on the value of θ_0 (and does not depend on the particular choice of θ_a).

- When a test obtained by Theorem 10.1 actually maximizes the power for every value of θ greater than θ0, it is said to be a ***uniformly most powerful test*** for $H_0: \theta = \theta_0$ versus $H_a: \theta > \theta_0$.

Example

EXAMPLE 10.23

Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and known variance σ^2. We wish to test $H_0: \mu = \mu_0$ against $H_a: \mu > \mu_0$ for a specified constant μ_0. Find the uniformly most powerful test with significance level α.

Example

SOLUTION 10.23

Consider the most powerful α-level test of $H_0: \mu = \mu_0$ versus $H_a^*: \mu = \mu_a$ for fixed $\mu_a > \mu_0$. Because $f(y|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right], y \in \mathbb{R}$, we have

$$L(\mu) = \prod_{i=1}^{n} f(y_i|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y-\mu)^2}{2\sigma^2}\right].$$

Because both H_0 and H_a^* are simple hypotheses, the most powerful test of H_0 versus H_a^* is given by

$$\frac{L(\mu_0)}{L(\mu_a)} < k \leftrightarrow \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y-\mu_0)^2}{2\sigma^2}\right]}{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(y-\mu_a)^2}{2\sigma^2}\right]} < k$$

Note that this is equivalent to $\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2 - \sum_{i=1}^{n}(y_i-\mu_a)^2\right]\right\} < k$.

Taking natural logarithms gives $-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2 - \sum_{i=1}^{n}(y_i-\mu_a)^2\right] < \ln(k)$

Example

SOLUTION 10.23

Taking natural logarithms gives

$$-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i-\mu_0)^2-\sum_{i=1}^n (y_i-\mu_a)^2\right]<\ln(k)$$
$$\to \sum_{i=1}^n (y_i-\mu_0)^2-\sum_{i=1}^n (y_i-\mu_a)^2>2\sigma^2\ln(k)$$
$$\to \sum_{i=1}^n y_i^2-2n\bar{y}\mu_0+n\mu_0^2-\sum_{i=1}^n y_i^2+2n\bar{y}\mu_a-n\mu_a^2>2\sigma^2\ln(k)$$
$$\to \bar{y}>\frac{-2\sigma^2\ln(k)-n\mu_0^2+n\mu_a^2}{2n(\mu_a-\mu_0)}=:k'(\because \mu_a>\mu_0)$$

Therefore, the most powerful test of H_0 versus H_a^* has rejection region given by $RR=\{\bar{y}>k'\}$.

Note that $\alpha=P(\bar{Y}\ in\ RR|\mu=\mu_0)=P(\bar{Y}>k'|\mu=\mu_0)=P\left(\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}}>\frac{k'-\mu_0}{\sigma/\sqrt{n}}\right)=P(Z>\sqrt{n}(k'-\mu_0)/\sigma)$.

Thus, $\sqrt{n}(k'-\mu_0)/\sigma=z_\alpha$ and $k'=\mu_0+z_\alpha\sigma/\sqrt{n}$ since Z follows a standard normal distribution under H_0.
Since neither $\bar{y}$ nor k' is a function of μ_a, this test is uniformly most powerful test for $H_0\!:\mu=\mu_0$ versus $H_a\!:\mu>\mu_0$ for a fixed α.

Motivation

- Theorem 10.1 provides a method of constructing most powerful tests for simple hypotheses when the distribution of the observations is known except for the value of a single unknown parameter.
- This method can sometimes be used to find uniformly most powerful tests for composite hypotheses that involve a single parameter.
- In many cases, the distribution of concern **has more than one unknown parameter**.
- In this section, we present a very general method that can be used to derive tests of hypotheses.
  - **The procedure works for simple or composite hypotheses and whether or not other parameters with unknown values are present.**

Motivation

- Suppose that a random sample is selected from a distribution and that the likelihood function $L(y_1, y_2, \dots, y_n | \theta_1, \theta_2, \dots, \theta_k)$ is a function of k parameters, $\theta_1, \theta_2, \dots, \theta_k$.

- To simplify notation, let $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ denote the vector of all k parameters and write

$$L(\Theta) = L(y_1, y_2, \dots, y_n | \theta_1, \theta_2, \dots, \theta_k)$$

- It may be the case that we are interested in testing hypotheses only about one of the parameters, say, θ_1

- For example, $\Theta = (\mu, \sigma^2)$. If we are interested in testing hypotheses about only the mean μ,
  - ✓ then σ^2—a parameter not of particular interest to us—is called a nuisance parameter.

- Thus, the likelihood function may be a function with both ***unknown nuisance parameters*** and a ***parameter of interest***.

Motivation

- ***Null hypothesis*** specifies that Θ (may be a vector) lies in a particular set of possible values—say, Ω_0
- ***Alternative hypothesis*** specifies that Θ lies in another set of possible values Ω_a, which does not overlap Ω_0.

$$\left.\begin{array}{|c|}\hline \Theta \in \Omega_0 \\ \hline \Theta \in \Omega_a \\ \hline \end{array}\right\} \Theta \in \Omega = (\Omega_0 \cup \Omega_a)$$

- For example, if we sample from a population with an exponential distribution with mean λ.
  - ✓ λ is the only parameter of the distribution, and $\Theta = \lambda$

- We are interested in testing

$$H_0\colon \lambda = \lambda_0 \text{ versus } H_a\colon \lambda \neq \lambda_0$$

  - ✓ $\Omega_0 = \{\lambda_0\}$
  - ✓ $\Omega_a = \{\lambda > 0\colon \lambda \neq \lambda_0\}$
  - ✓ $\Omega = \Omega_0 \cup \Omega_a = \{\lambda_0\} \cup \{\lambda > 0\colon \lambda \neq \lambda_0\} = \{\lambda\colon \lambda > 0\}$

- Either or both of the hypotheses H_0 and H_a can be composite because they might contain multiple values of the parameter of interest or because other unknown parameters may be present.

Likelihood Ratio Test

A Likelihood Ratio Test

Define λ by

$$\lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)}$$

A likelihood ratio test of $H_0: \Theta \in \Omega_0$ versus $H_a: \Theta \in \Omega_a$ employs λ as a test statistics, and the rejection region is determined by $\lambda \leq k$

- If $L(\widehat{\Omega}_0) = L(\widehat{\Omega})$, then a best explanation for the observed data can be found inside Ω_0, and we should not reject the null hypothesis $H_0: \Theta \in \Omega_0$.
- However, If $L(\widehat{\Omega}_0) < L(\widehat{\Omega})$ then the best explanation for the observed data can be found inside Ω_a, and we should consider rejecting H_0 in favor of H_a.
- A likelihood ratio test is based on the ratio $L(\widehat{\Omega}_0)/L(\widehat{\Omega})$
- It can be shown that $0 \leq \lambda \leq 1$. A value of λ close to zero indicates that the likelihood of the sample is much smaller under H_0 than it is under H_a.
  - Therefore, the data suggest favoring H_a over H_0.

Example

EXAMPLE 10.24

Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. We want to test $H_0: \mu = \mu_0$ versus $H_a: \mu > \mu_0$. Find the appropriate likelihood ratio test.

Example

SOLUTION 10.24

In this case, $\Theta = (\mu, \sigma^2)$. Notice that $\Omega_0 = \{(\mu_0, \sigma^2): \sigma^2 > 0\}, \Omega_a = \{(\mu, \sigma^2): \mu > \mu_0, \sigma^2 > 0\}$, and hence that $\Omega = \{(\mu, \sigma^2): \mu \geq \mu_0, \sigma^2 > 0\}$.

For normal distribution, we have $L(\Theta) = L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[-\sum_{i=1}^n \frac{(y_i-\mu)^2}{2\sigma^2}\right]$.

From Example 9.15,we see that $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \mu_0)^2$ when $\mu = \mu_0$.

Thus, $L(\widehat{\Omega}_0) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0{}^2}\right)^{n/2} \exp\left[-\sum_{i=1}^n \frac{(y_i-\mu_0)^2}{2\hat{\sigma}_0{}^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0{}^2}\right)^{n/2} e^{-n/2}$.

Note that $\ln\left(L(\mu, \sigma^2)\right) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2$ and $\ln(L(\mu, \sigma^2))$ is maximized when $\mu = \bar{y}$.

For $\mu > \bar{y}, \partial L(\mu, \sigma^2)/\partial\mu < 0$ so $L(\mu, \sigma^2)$ decreases for $\mu > \bar{y}$. Thus, over the set $\Omega = \{(\mu, \sigma^2): \mu \geq$

Example

SOLUTION 10.24

Just as earlier, the value of $\sigma^2 \in \Omega$ that maximizes $L(\mu, \sigma^2)$ is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$, which yields

$$L(\widehat{\Omega}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n} \frac{(y_i-\mu_0)^2}{2\hat{\sigma}^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} e^{-n/2}.$$

Thus, $\lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) = \begin{cases} \left[\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\mu_0)^2}\right]^{n/2}, & if\ \bar{y} \geq \mu_0 \\ 1, & if\ \bar{y} < \mu_0 \end{cases}.$

Since λ is always less than or equal to 1, we need only to focus on when $\bar{y} \geq \mu_0$.

Because $\sum_{i=1}^{n}(y_i - \mu_0)^2 = \sum_{i=1}^{n}[(y_i - \bar{y}) + (\bar{y} - \mu_0)]^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu_0)^2$, the rejection region $\lambda \leq k < 1$ is equivalent to

$$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\mu_0)^2} < k^{2/n} := k' \rightarrow \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu_0)^2} < k' \rightarrow \frac{1}{1+\frac{n(\bar{y}-\mu_0)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} < k'$$

Example

SOLUTION 10.24

This inequality is equivalent to

$$\frac{n(\bar{y}-\mu_0)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} > \frac{1}{k'} - 1 =: k'' \rightarrow \frac{n(\bar{y}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2} > (n-1)k'' \rightarrow \frac{\sqrt{n}(\bar{y}-\mu_0)}{s} > \sqrt{(n-1)k''}(\because \bar{y} \geq \mu_0)$$

where $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$.

Notice that $\frac{\sqrt{n}(\bar{y}-\mu_0)}{s}$ is a t statistic employed in previous sections. Consequently, the likelihood ratio test is equivalent to the t test of Section 10.8.

Finding rejection region for likelihood test

- Situations in which the likelihood ratio test assumes a well-known form are not uncommon.
- For most practical problems, the likelihood ratio method produces the best possible test, in terms of power.
- Unfortunately, the likelihood ratio method does not always produce a test statistic with a known probability distribution
- If the samplesizeislarge,however,wecanobtainanapproximationtothedistributionof λ if some reasonable "regularity conditions" are satisfied by the underlying population distribution(s).
  - ✓ These are general conditions that hold for most (but not all) of the distributions that we have considered.
  - ✓ The regularity conditions mainly involve the existence of derivatives, with respect to the parameters, of the likelihood function.

Definition

THEOREM 10.2

Let $Y_1, Y_2, \ldots, Y_n$ have joint likelihood function $L(\Theta)$. Let r_0 denote the number of free parameters that are specified by $H_0 : \Theta \in \Omega_0$ and let r denote the number of free parameters (nuisance parameter, that is unknown) specified by the statement $\Theta \in \Omega$. Then, for large n, $-2\ln(\lambda)$ has approximately a χ^2 distribution with $r_0 - r$ df.

- Theorem 10.2 allows us to use the table of the χ^2 distribution to find rejection regions with fixed α when n is large.
- Because the likelihood ratio test specifies that we use RR: $\{\lambda < k\}$, this rejection may be re written as RR: $\{-2\ln(\lambda) > -2\ln(k) = k^*\} = \{\chi^2 > \chi^2_\alpha\}$
- For large sample sizes, if we desire an α −level test, Theorem 10.2 implies that $k^* \approx \chi^2_\alpha$. That is, a large-sample likelihood ratio test has rejection region given by

$$-2\ln(\lambda) > \chi^2_\alpha \text{, where } \chi^2_\alpha \text{ is based on } r_0 - r \text{ df}$$

- It is important to realize that large-sample likelihood ratio tests are based on $-2\ln(\lambda)$ where λ is the original likelihood ratio, $\lambda = L(\widehat{\Omega}_0)/L(\widehat{\Omega})$

Example

EXAMPLE 10.25

Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a manufacturing plant. One hundred independent observations on the number of complaints gave means $\bar{x} = 20$ for shift 1 and $\bar{y} = 22$ for shift 2. Assume that the number of complaints per week on the ith shift has a Poisson distribution with mean θ_i , for $i = 1, 2$. Use the likelihood ratio method to test $H_0\colon\ \theta_1 = \theta_2$ versus $H_a\colon\ \theta_1 \neq \theta_2$ with $\alpha \approx .01$.

Example

SOLUTION 10.25

The joint likelihood of the sample is given by $L(\theta_1, \theta_2) = \frac{1}{\prod_{i=1}^{n} x_i! \prod_{j=1}^{n} y_j!} \theta_1^{\sum_i x_i} e^{-n\theta_1} \theta_2^{\sum_j y_j} e^{-n\theta_2}$.

In this example, $\Theta = (\theta_1, \theta_2)$ and $\Omega_0 = \{(\theta_1, \theta_2): \theta_1 = \theta_2 = \theta\}$, where θ is unknown. Hence, the likelihood function under H_0 is $L(\theta) = \frac{1}{\prod_{i=1}^{n} x_i! \prod_{j=1}^{n} y_j!} \theta^{\sum_i x_i + \sum_j y_j} e^{-2n\theta}$

Notice that for $\Theta \in \Omega_0$, $L(\theta)$ is maximized when $\hat{\theta} = \frac{1}{2n}\left(\sum_i x_i + \sum_j y_j\right) = \frac{1}{2}(\bar{x} + \bar{y})$.

Also, $\Omega_a = \{(\theta_1, \theta_2): \theta_1 \neq \theta_2\}$ and $\Omega = \{(\theta_1, \theta_2): \theta_1, \theta_2 > 0\}$. For $\Theta \in \Omega$, it is easy to verify that $L(\theta_1, \theta_2)$ is maximized at $\hat{\theta}_1 = \bar{x}$ and $\hat{\theta}_2 = \bar{y}$.

Thus, $\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{\frac{1}{\prod_{i=1}^{n} x_i! \prod_{j=1}^{n} y_j!} \hat{\theta}^{n\bar{x}+n\bar{y}} e^{-2n\hat{\theta}}}{\frac{1}{\prod_{i=1}^{n} x_i! \prod_{j=1}^{n} y_j!} \hat{\theta}_1^{n\bar{x}} e^{-n\hat{\theta}_1} \hat{\theta}_2^{n\bar{y}} e^{-n\hat{\theta}_2}} = \frac{(\hat{\theta})^{n\bar{x}+n\bar{y}}}{(\bar{x})^{n\bar{x}} (\bar{y})^{n\bar{y}}} = \frac{21^{100(20+22)}}{20^{100(20)} 22^{100(22)}}$

and hence $-2\ln(\lambda) = -(2)[4200\ln(21) - 2000\ln(20) - 2200\ln(22)] = 9.53$.

In this application, the number of free parameters in Ω is $k = 2$, and that of fixed parameters is $r = 0$. In Ω_0, $r_0 = 1$ of free parameters is fixed. Theorem 10.2 says that $-2\ln(\lambda)$ has an approximately χ^2 distribution with $r_0 - r = 1 - 0 = 1$ df. Since $-2\ln(\lambda) > \chi^2_{.01} = 6.635$, we reject H_0.