Chapter 9: Tests of Hypotheses for a Single Sample

LEARNING OBJECTIVES

1. Introduction
2. Test on the μ of NORMDIST
   - σ^2 known; σ^2 unknown
3. Test on the σ^2 : Normal Distribution
4. Test on the *p*: Large-sample

Definition — Statistical Hypotheses

A statistical hypothesis is a statement about the parameters of one or more populations.

Example

Suppose that we are interested in the burning rate of a solid propellant used to power aircrew escape systems.

- Burning rate is a random variable
- We are interested in deciding whether or not the mean burning rate is 50 centimeters per second.

Two-sided Alternative Hypothesis

$$H_0:\ \mu = 50;\ H_1:\ \mu \neq 50$$

Null hypothesis

Alternative hypothesis

One-sided Alternative Hypotheses

$$H_0:\ \mu = 50;\ H_1:\ \mu > 50$$

$$H_0:\ \mu = 50;\ H_1:\ \mu < 50$$

Test of a Hypothesis

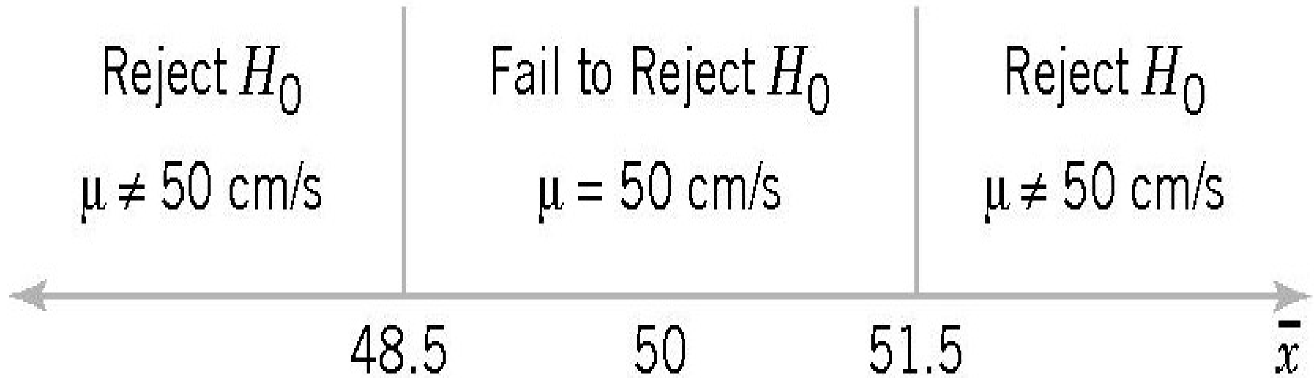
- A procedure leading to a decision about a particular hypothesis

- Hypothesis-testing procedures rely on using the information in a random sample from the population of interest.

- If this information is *consistent* with the hypothesis, then we will conclude that the hypothesis is true; if this information is *inconsistent* with the hypothesis, we will conclude that the hypothesis is false.

Example 1

$$H_0: \mu = 50;\ H_1: \mu \neq 50$$



Decision criteria for testing H_0: $\mu = 50$; H_1: $\mu \neq 50$

Table 9-1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	no error	type II error
Reject H_0	type I error	no error

$\alpha = P(\text{ type I error }) = P(\text{ reject } H_0 \text{ when } H_0 \text{ is true })$

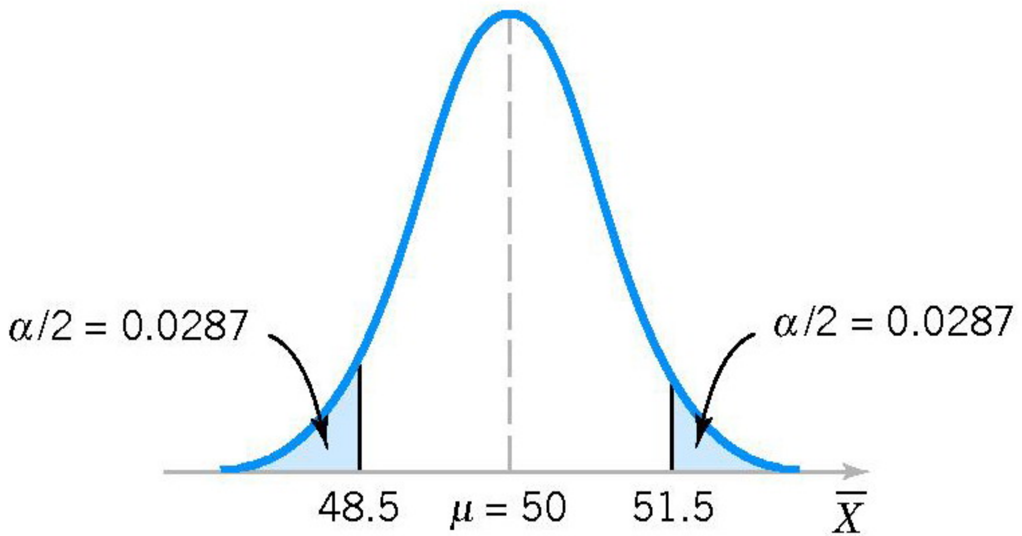
Significance level

$\beta = P(\text{ type II error}) = P(\text{ fail to reject } H_0 \text{ when } H_0 \text{ is false})$

To calculate α, we assume that σ = 2.5 and n = 10.

$$\alpha = P(\overline{X} < 48.5 \ when \ \ \mu = 50) + P(\overline{X} > 51.5 \ when \ \ \mu = 50)$$
$$= P(Z < -1.90) + P(Z > 1.90) \simeq 0.0574$$

$$\boxed{= \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\overline{X} - 50}{2.5 / \sqrt{10}}}$$



ERRORS

$$\beta = P(48.5 < \overline{X} < 51.5 \ \ when \ \ \mu = 52)$$
$$= P(-4.43 < Z < -0.63) \simeq 0.2643$$

$$\beta = P(48.5 < \overline{X} < 51.5 \ \ when \ \ \mu = 50.5)$$
$$= P(-2.53 < Z < 1.27) \simeq 0.8923$$

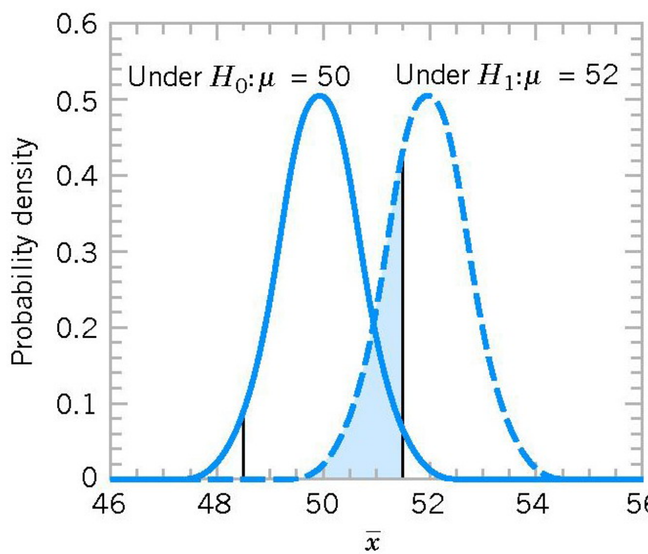


Figure 9-3 The probability of type II error when μ = 52 and n = 10.

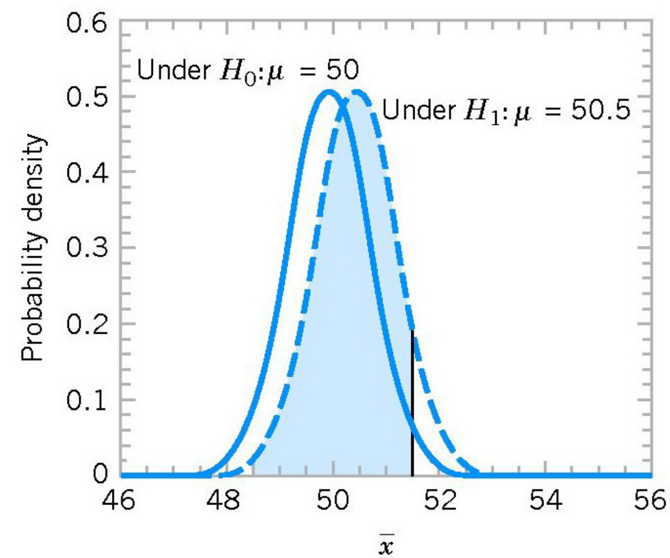


Figure 9-4 The probability of type II error when μ = 50.5 and n = 10.

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	$\boxed{0.5000}$	$\boxed{0.9705}$
$48.5 < \bar{x} < 51.5$	16	0.0164	0.2119	$\boxed{0.9445}$
$48 < \bar{x} < 52$	16	$\boxed{0.0014}$	$\boxed{0.5000}$	$\boxed{0.9918}$

1. The size of the critical region, and consequently the probability of a type I error , can always be reduced by appropriate selection of the critical values.

2. Type I and type II errors are related: β ↓ then α ↑ and α ↓ then β ↑ , provided that *n* fix.

3. An increase in sample size will generally reduce both α and β, provided that the critical values are held constant.

4. When the null H_0 is false, β ↑ as the true value of the parameter approaches the value hypothesized in H_0,β ↓ as the difference between the true mean and the hypothesized value increases.

- Because we can control the probability of making a type I error (or significance level), a logical question is what value should be used.
- The type I error probability is a measure of risk, specifically, the risk of concluding that the null hypothesis is false when it really isn't.
- scientific and engineering practice is to use the value, **significance level is 0.05** in most situations.
- In the rocket propellant problem with , this would correspond to critical values of 48.45 and 51.55

Definition power of a statistical test

- Note that, on the other hand, the probability of type II error is not a constant, but depends on the true value of the parameter.
- The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

• The power (hiệu năng) is computed as 1 - beta, and power can be interpreted as ***the probability of correctly rejecting a false null hypothesis.***

• We often compare statistical tests by comparing their power properties.

Definition power of a statistical test

- We often compare statistical tests by comparing their power properties.
- Power is a very descriptive and concise measure of the **sensitivity** of a statistical test, mean that the ability of the test to detect differences.

Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

That is, this test will correctly reject and "detect" this difference 73.57% of the time. If this value of power is judged to be too low, the analyst can increase either alpha or the sample size *n*.

General Procedure for Hypothesis Tests

1. **Parameter of interest:** From the problem context, identify the parameter of interest.
2. **Null hypothesis, H_0:** State the null hypothesis, H_0.
3. **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1.
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject H_0 if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Hypothesis-testing problems

Two-sided test: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$

One-sided test: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$

H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$

TEST ON THE μ OF NORMDIST: σ^2 KNOWN

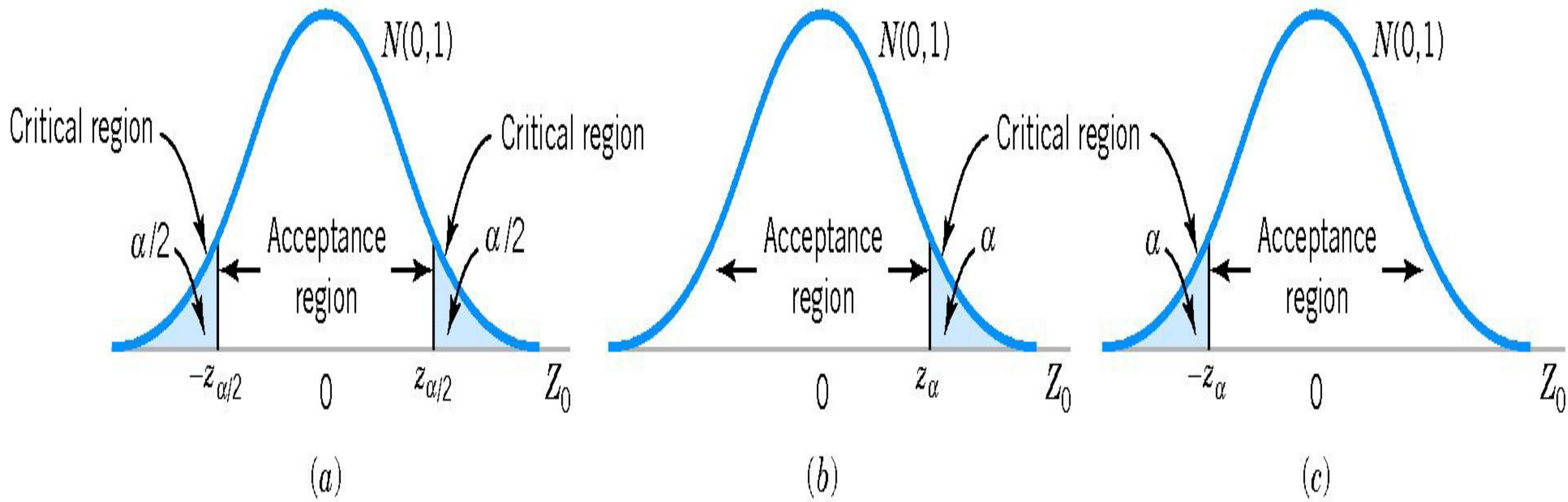


Figure 9-6 The distribution of Z_0 when $H_0: \mu = \mu_0$ is true, with critical region for (a) the two-sided alternative $H_1: \mu \neq \mu_0$, (b) the one-sided alternative $H_1: \mu > \mu_0$, and (c) the one-sided alternative $H_1: \mu < \mu_0$.

Hypothesis Tests on the Mean — z-test

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic: $$Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Alternative hypothesis	Rejection criteria
H_1: $\mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
H_1: $\mu > \mu_0$	$z_0 > z_\alpha$
H_1: $\mu < \mu_0$	$z_0 < -z_\alpha$

Example 2

A melting point test of n = 10 samples of a binder used in manufacturing a rocket propellant resulted in $\bar{x} = 154.2^0$. Assume that melting point is normally distributed with σ = 1.5^0.

Test H_0: μ = 155; H_1: μ ≠ 155 using α = 0.05.

Solution: Test statistic:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{154.2 - 155}{1.5 / \sqrt{10}} \simeq -1.69$$

$z_{\alpha/2} = z_{0.025} = 1.96.$

$|z_0| < z_{\alpha/2}.$

Fail to reject H_0 at α = 0.05

Example 3

9-5. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a standard deviation of 0.5 kilograms. The company wishes to test the hypothesis H_0: $\mu = 12$ against H_1: $\mu < 12$, using a random sample of four specimens.

(a) What is the type I error probability if the critical region is defined as $\bar{x} < 11.5$ kilograms?

(b) Find β for the case where the true mean elongation is 11.25 kilograms.

(c) Find β for the case where the true mean is 11.5 kilograms.

9-6. Repeat Exercise 9-5 using a sample size of $n = 16$ and the same critical region.

P-Values in Hypothesis Tests

Definition

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

- The *P*-value is the probability that the test statistic will take on a value that is at least as extreme as the observed value of the statistic when the null hypothesis *H0* is true.
- *P*-value conveys much information about the weight of evidence against *H0*, and so a decision maker can draw a conclusion at *any* specified level of significance.

P-Values in Hypothesis Tests

Definition

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

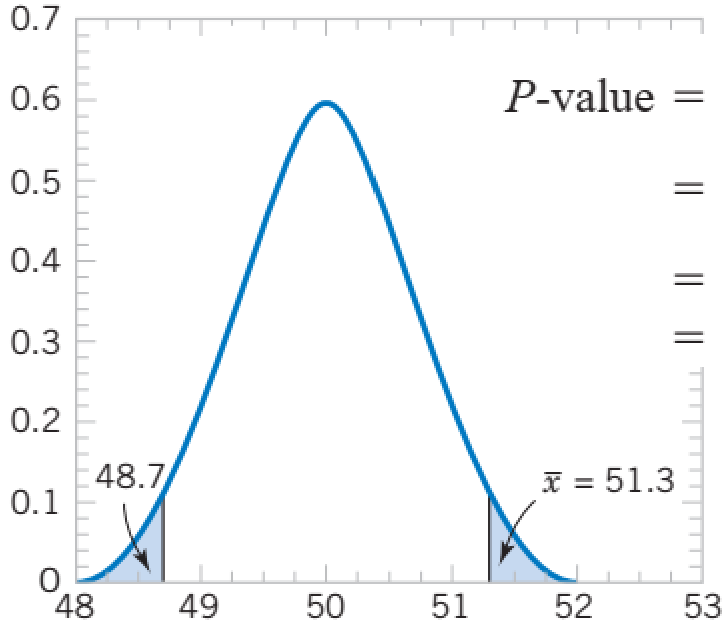
$$P = \begin{cases} 2[1 - \Phi(|z_0|)] & \text{for a two-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu \neq \mu_0 \\ 1 - \Phi(z_0) & \text{for a upper-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu > \mu_0 \\ \Phi(z_0) & \text{for a lower-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu < \mu_0 \end{cases}$$

P-Values in Hypothesis Tests

Consider the two-sided hypothesis test for burning rate

$$H_0\colon \mu = 50 \quad H_1\colon \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is $\bar{x} = 51.3$ centimeters



$$\begin{aligned} P\text{-value} &= 1 - P(48.7 < \bar{X} < 51.3) \\ &= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right) \\ &= 1 - P(-2.08 < Z < 2.08) \\ &= 1 - 0.962 = 0.038 \end{aligned}$$

P-Values in Hypothesis Tests

Consider the two-sided hypothesis test for burning rate

$$H_0: \mu = 50 \quad H_1: \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is $\bar{x} = 51.3$ centimeters

- Compared to the "standard" level of significance 0.05, our observed *P*-value is smaller, so if we were using a fixed significance level of 0.05, the null hypothesis would be rejected.
- In fact, the null hypothesis *H0* would be rejected at *any* level of significance greater than or equal to 0.038.
- The *P*-value is the smallest level of significance that would lead to rejection of H0

TEST ON THE μ OF NORMDIST: σ^2 KNOWN

Introduction

Test on the μ of NORMDIST

σ^2 known

σ^2 unknown

Test on the p

Summary

19/05/2021

Testing Hypotheses on the Mean, Variance Known (*Z*-Tests)

Null hypothesis: $H_0\colon \mu = \mu_0$

Test statistic: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

Alternative Hypotheses	*P*-Value	Rejection Criterion for Fixed-Level Tests
$H_1\colon \mu \neq \mu_0$	Probability above $\lvert z_0\rvert$ and probability below $-\lvert z_0\rvert$, $P = 2[1 - \Phi(\lvert z_0\rvert)]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1\colon \mu > \mu_0$	Probability above z_0, $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1\colon \mu < \mu_0$	Probability below z_0, $P = \Phi(z_0)$	$z_0 < -z_\alpha$

The *P*-values and critical regions for these situations are shown in Figs. 9-7 and 9-8.

Type II Error and Choice of Sample Size

Finding the Probability of Type II Error β

Consider the two-sided hypotheses

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The test statistic Z_0 is

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

Therefore, the distribution of Z_0 when H_1 is true is

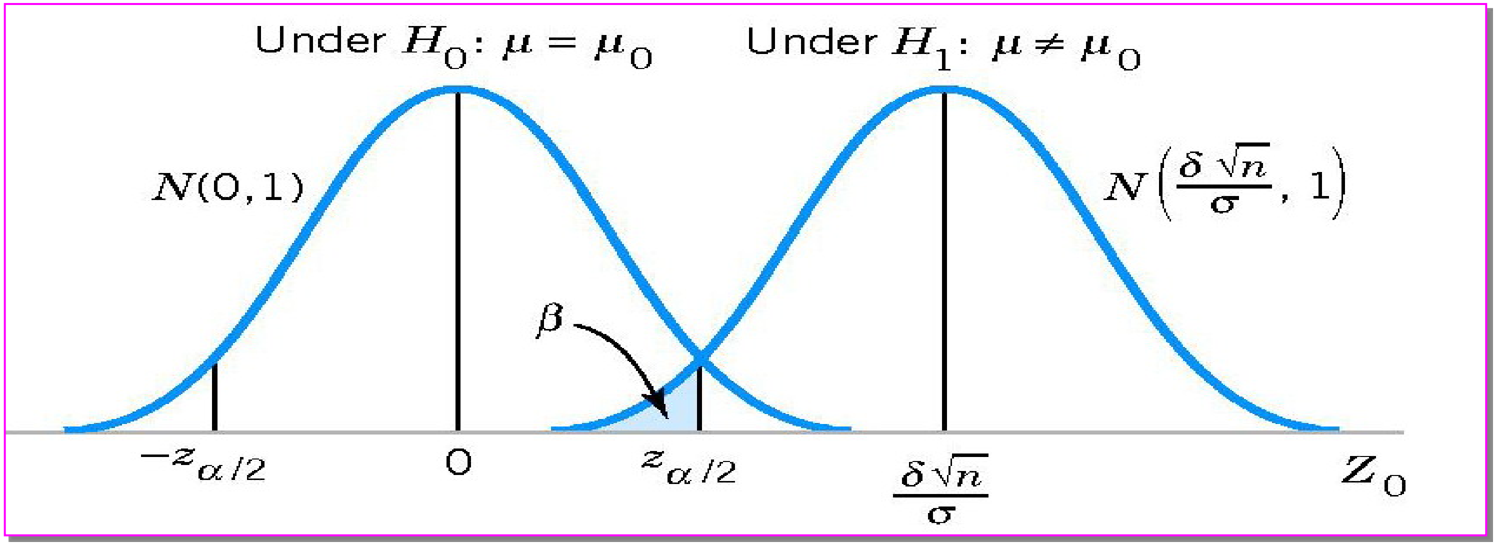
$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \tag{9-19}$$

Type II Error and Choice of Sample Size

Probability of type II error β for a two-sided test

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

where $\delta = \mu - \mu_0$



Type II Error and Choice of Sample Size

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

or, if $\delta > 0$,

$$\beta \simeq \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \tag{9-21}$$

since $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) \simeq 0$ when δ is positive. Let z_β be the 100β upper percentile of the standard normal distribution. Then, $\beta = \Phi(-z_\beta)$. From Equation 9-21,

$$-z_\beta \simeq z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}$$

Type II Error and Choice of Sample Size

Sample Size Formulas

Two-sided test

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2}$$

One-sided test

$$n \simeq \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\delta^2}$$

where $\delta = \mu - \mu_0$.

Exercise

The life in hours of a battery is known to be approximately normally distributed, with standard deviation σ =1.25 hours. A random sample of 10 batteries has a mean life of =40.5 hours.

(a) Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.05$.

(b) What is the *P*-value for the test in part (a)?

(c) What is the β-error for the test in part (a) if the true mean life is 42 hours?

(d) What sample size would be required to ensure that does not exceed 0.10 if the true mean life is 44 hours?

Hypothesis Tests on the Mean — *t*-test

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic: $$T_0 = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$

Alternative hypothesis	Rejection criteria
H_1: $\mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\alpha/2,\, n-1}$
H_1: $\mu > \mu_0$	$t_0 > t_{\alpha,\, n-1}$
H_1: $\mu < \mu_0$	$t_0 < - t_{\alpha,\, n-1}$

TEST ON THE μ OF NORMDIST: σ^2 UNKNOWN

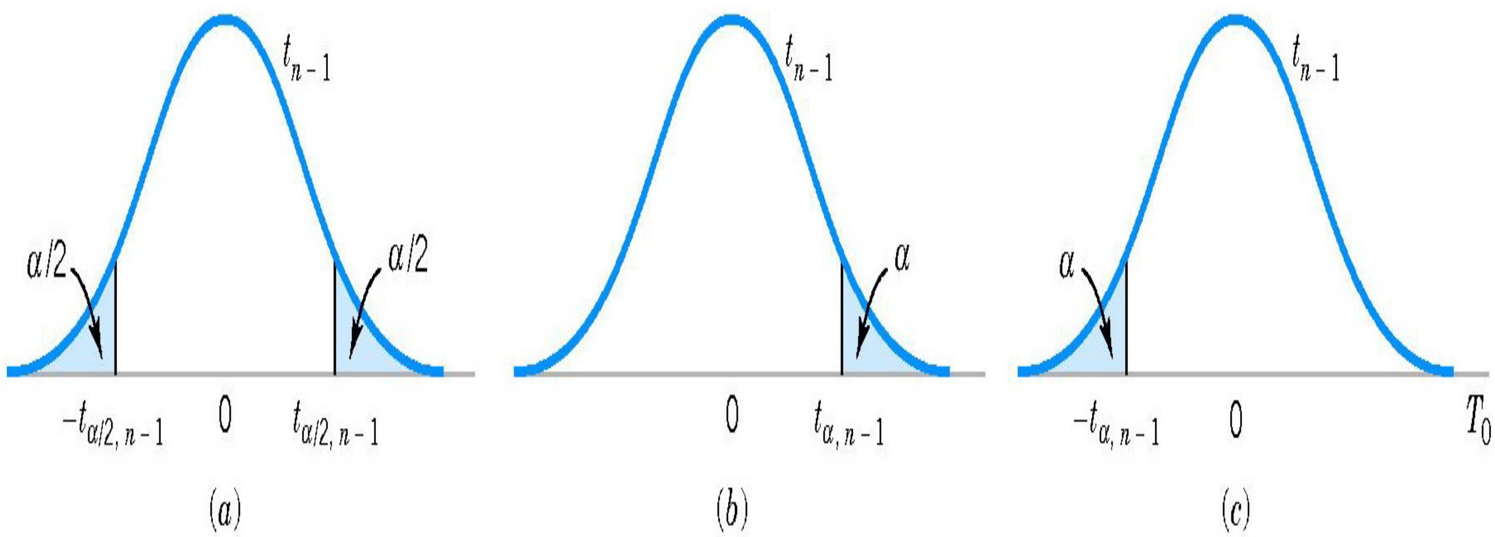


Figure 9-8 The reference distribution for H_0: $\mu = \mu_0$ with critical region for (a) H_1: $\mu \neq \mu_0$, (b) H_1: $\mu > \mu_0$, and (c) H_1: $\mu < \mu_0$.

Exercise

Consider the dissolved oxygen concentration at TVA dams. The observations are (in milligrams per liter): 5.0, 3.4, 3.9, 1.3, 0.2, 0.9, 2.7, 3.7, 3.8, 4.1, 1.0, 1.0, 0.8, 0.4, 3.8, 4.5, 5.3, 6.1, 6.9, and 6.5.

(a) Test the hypotheses H_0: $\mu = 4$; H_1: $\mu \neq 4$. Use $\alpha = 0.01$.

(b) What is the P-value in part (a)?

(c) Compute the power of the test if the true mean $\mu = 3$.

Hypothesis Tests on the Mean

Null hypothesis: $H_0: p = p_0$

Test statistic: $$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

Alternative hypothesis	Rejection criteria
$H_1: p \neq p_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
$H_1: p > p_0$	$z_0 > z_\alpha$
$H_1: p < p_0$	$z_0 < -z_\alpha$

Example

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish roughness that exceeds the specifications. Does this data present strong evidence that the proportion of crankshaft bearings exhibiting excess surface roughness exceeds 0.10? State and test the appropriate hypotheses using α = 0.05.

Solution: H_0: $p = 0.10$; H_1: $p > 0.10$

=1.65

$$\hat{p} = \frac{10}{85} \simeq 0.12 \Rightarrow z_0 = \frac{0.12 - 0.10}{\sqrt{\dfrac{0.10(1-0.10)}{85}}} = 0.61 < z_\alpha$$

Fail to reject H_0 at α = 0.05.

Type II Error and Choice of Sample Size

Probability of type II error β

Two-sided test

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

One-sided test: $p > p_0$

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

One-sided test: $p < p_0$

$$\beta = 1 - \Phi\left(\frac{p_0 - p - z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

Type II Error and Choice of Sample Size

Sample Size Formulas

Two-sided test

$$n = \left[\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2$$

One-sided test

$$n = \left[\frac{z_{\alpha}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2$$

Hypothesis Tests on Variance

Null hypothesis: $H_0: \sigma^2 = \sigma_0^2$

Test statistic: $\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

Alternative hypothesis	Rejection criteria
$H_1: \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or $\chi_0^2 < -\chi^2_{\alpha/2,n-1}$
$H_1: \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha,n-1}$
$H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 < -\chi^2_{\alpha,n-1}$

EXAMPLE 9-8 Automated Filling

An automated filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

Using the seven-step procedure results in the following:

1. **Parameter of Interest:** The parameter of interest is the population variance σ^2.
2. **Null hypothesis:** $H_0: \sigma^2 = 0.01$
3. **Alternative hypothesis:** $H_1: \sigma^2 > 0.01$

EXAMPLE 9-8 Automated Filling

Using the seven-step procedure results in the following:

1. **Parameter of Interest:** The parameter of interest is the population variance σ^2.
2. **Null hypothesis:** H_0: $\sigma^2 = 0.01$
3. **Alternative hypothesis:** H_1: $\sigma^2 > 0.01$
4. **Test statistic:** The test statistic is

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

5. **Reject H_0:** Use $\alpha = 0.05$, and reject H_0 if $\chi_0^2 > \chi_{0.05,19}^2 = 30.14$.
6. **Computations:**

$$\chi_0^2 = \frac{19(0.0153)}{0.01} = 29.07$$

7. **Conclusions:** Since $\chi_0^2 = 29.07 < \chi_{0.05,19}^2 = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2. So there is no strong evidence of a problem with incorrectly filled bottles.

SUMMARY

We have studied:

1. Test on the μ of NORMDIST

- σ^2 known
- σ^2 unknown

2. Test on the σ^2 : Normal Distribution

3. Test on the *p*: Large-sample