# Chapter 8. Tests of Hypotheses Based on a Single Sample

Weiqi Luo (骆伟祺)
School of Data & Computer Science
Sun Yat-Sen University
Email：luoweiqi@mail.sysu.edu.cn

# Chapter 8: Tests of Hypotheses Based on a Single Sample

- 8.1. Hypotheses and Test Procedures
- 8.2. Tests About a Population Mean
- 8.3. Tests Concerning a Population Proportion
- 8.4. P-Values
- 8.5. Some Comments on Selecting a Test Procedure

# 8.1. Hypotheses and Test Procedures

- A parameter can be estimated from sample data either by a single number (Chap 6. a point estimate) or an entire interval of plausible values (Chap 7. a confidence interval).

- However, the objective of an investigation is not to estimate a parameter but to decide which of two contradictory claims about the parameter is correct.

- Methods for accomplishing this comprise the part of statistical inference called *hypothesis testing*

# 8.1. Hypotheses and Test Procedures

- A **statistical hypothesis,** or just *hypothesis*, is a claim or assertion either about the value of a single parameter (population characteristic or characteristic of a probability distribution), about the values of several parameters, or about the form of an entire probability distribution.

- In any hypothesis-testing problem, there are two contradictory hypotheses under consideration.

- The objective is to decide, based on sample information, which of the two hypotheses is correct

# 8.1. Hypotheses and Test Procedures

- The **null hypothesis,** denoted by $H_0$, is the claim that is initially assumed to be true. The **alternative hypothesis,** denoted by $H_a$, is the assertion that is contradictory to $H_0$.

- The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that $H_0$ is false. If the sample does not strongly contradict $H_0$, we will continue to believe in the plausibility of the null hypothesis.

- The two possible conclusions from a hypothesis-testing analysis are then *reject* $H_0$ or *fail to reject* $H_0$.

# 8.1. Hypotheses and Test Procedures

- In our treatment of hypothesis testing, $H_0$ will generally be stated as an equality claim, e.g., $H_0$: $\theta = \theta_0$

Let p denote the true proportion of defective boards resulting from the changed process. The suggested alternative hypothesis was $H_a$: $p<0.10$ , the claim that the defective rate is reduced by the process modification. A natural choice of $H_0$ in this situation is the claim that $H_0$: $p \geq 0.10$ according to which the new process is either no better or worse than the one currently used.

We will instead consider $H_0$: $p = 0.10$ vs. $H_a$: $p<0.10$ .

The rationale for using this simplified null hypothesis is that any reasonable decision procedure for deciding between $H_0$ and $H_a$ will also be reasonable for deciding between $p\geq 0.10$ and $H_a$

# 8.1. Hypotheses and Test Procedures

- The alternative to the null hypothesis $H_0$: $\theta = \theta_0$ will look like one of the following three assertions:

1. $H_a$: $\theta > \theta_0$ (in which case the implicit null hypothesis is $\theta \leq \theta_0$),
2. $H_a$: $\theta < \theta_0$ (in which case the implicit null hypothesis is $\theta \geq \theta_0$), or
3. $H_a$: $\theta \neq \theta_0$

The number $\theta_0$ that appears in both $H_0$ and $H_a$ (separates the alternative from the null) is called the **null value.**

# 8.1. Hypotheses and Test Procedures

- A test of $H_0$: $p = 0.1$ versus $H_a$: $p < 0.1$ in the circuit board problem might be based on examining a random sample of $n=200$ boards.

Let X denote the number of defective boards in the sample, a binomial random variable; x represents the observed value of X. If $H_0$ is true, $E(x) = np = 20$, whereas we can expect fewer than 20 defective boards if $H_a$ is true. A value x just a bit below 20 does not strongly contradict $H_0$, so it is reasonable to reject $H_0$ only if x is substantially less than 20. one such test procedure is to reject $H_0$ if $x \leq 15$ and not reject $H_0$ otherwise.

# 8.1. Hypotheses and Test Procedures

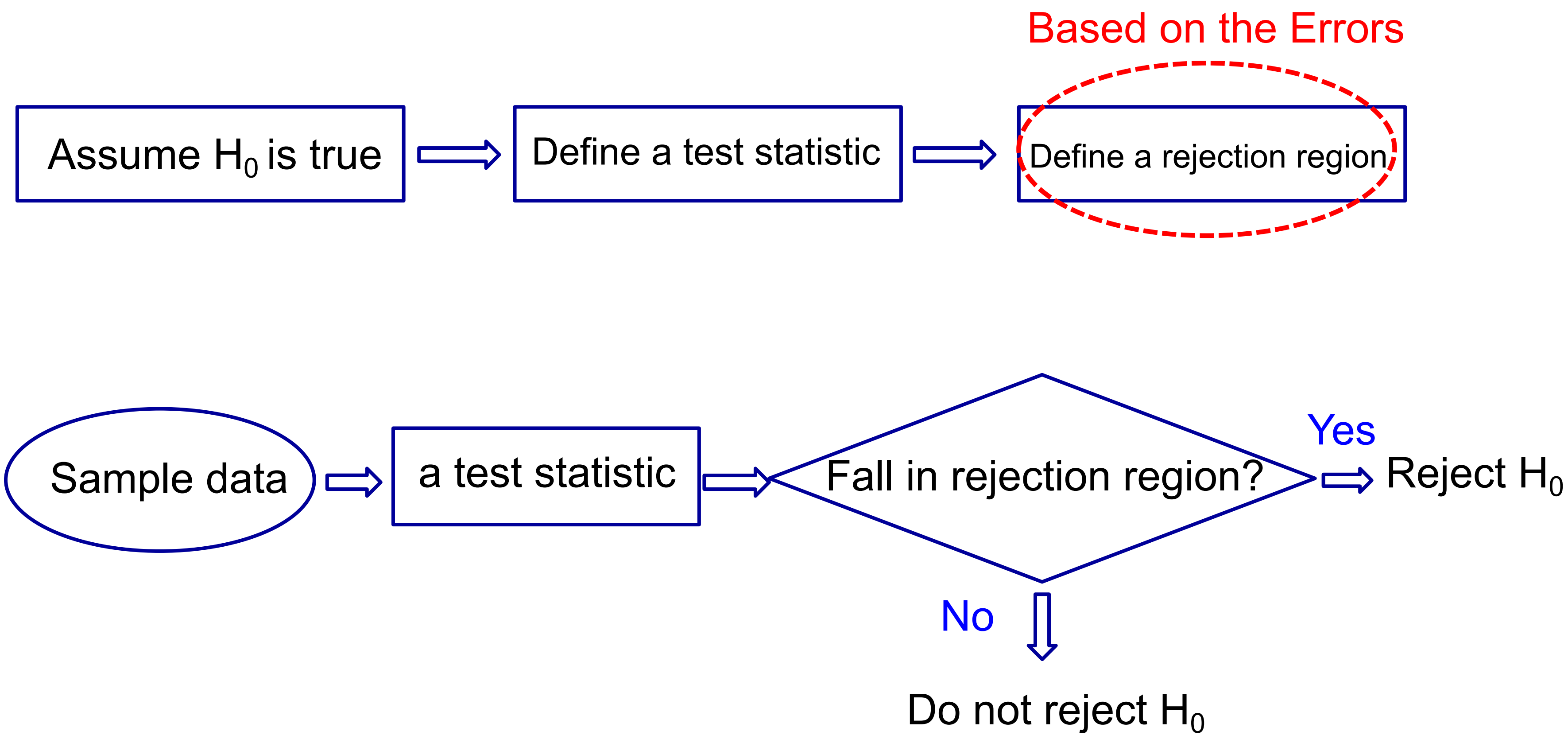
- A **test of hypotheses** is a method for using sample data to decide whether the null hypothesis should be rejected.

A test procedure is specified by the following:

1. A **test statistic,** a function of the sample data on which the decision (reject $H_0$ or do not reject $H_0$) is to be based
2. A **rejection region,** the set of all test statistic values for which $H_0$ will be rejected

The null hypothesis will then be rejected if and only if the observed or computed test statistic value falls in the rejection region.

# 8.1. Hypotheses and Test Procedures



# 8.1. Hypotheses and Test Procedures

## Two types of Errors in Hypothesis Testing

A **type I error** consists of rejecting the null hypothesis $H_0$ when it is true. A **type II error** involves not rejecting $H_0$ when $H_0$ is false.

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Reject $H_0$ | Type I error $\alpha$ | |
| Do not reject $H_0$ | | Type II error $\beta$ |

A good procedure is one for which the probability of making either type of error is small.

Because $H_0$ specifies a unique value of the parameter, there is a single value of $\alpha$. However, there is a different value of $\beta$ for each value of the parameter consistent with $H_a$. The choice of a particular rejection region cutoff value fixes the probabilities of type I and type II errors.

- Example 8.1

A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed in an effort to increase this percentage. Let $p$ denote the proportion of all 10-mph crashes with this new bumper that result in no visible damage. The hypotheses to be tested are $H_0$: p=0.25 (no improvement) versus $H_a$: p>0.25 .

The test will be based on an experiment involving n=20 independent crashes with prototypes of the new design. Intuitively, $H_0$ should be rejected if a substantial number of the crashes show no damage.

- Example 8.1 (Cont')

Test statistic: $X$ = the number of crashes with no visible damage

Rejection region: $R_8 = \{8,9,10,\ldots,19,20\}$; that is reject $H_0$ if $x>=8$, where x is the observed value of the test statistic.

When $H_0$ is true, $X$ has a binomial probability distribution with n=20, p = 0.25

$$\begin{aligned}\alpha &= P(\text{type I error}) = P(H_0 \text{ is rejected when it is true}) \\ &= P(X \geq 8 \text{ when } X \sim \text{Bin}(20, .25)) = 1 - B(7; 20, .25) \\ &= 1 - .898 = .102\end{aligned}$$

That is, when $H_0$ is actually true, roughly 10% of all experiments consisting of 20 crashes would result in $H_0$ being incorrectly rejected (a type I error).

- Example 8.1 (Cont')

In contrast to α, there is not a single β. Instead, there is a different β for each different *p* that exceeds 0.25. For instance

$$\begin{aligned}\beta(.3) &= P(\text{type II error when } p = .3)\\ &= P(H_0 \text{ is not rejected when it is false because } p = .3)\\ &= P(X \le 7 \text{ when } X \sim \text{Bin}(20, .3)) = B(7; 20, .3) = .772\end{aligned}$$

When *p* is 0.3 rather than 0.25, roughly 77% of all experiments of this type would result in $H_0$ being incorrectly not rejected!

| $p$ | .3 | .4 | .5 | .6 | .7 | .8 |
|---|---|---|---|---|---|---|
| $\beta(p)$ | .772 | .416 | .132 | .021 | .001 | .000 |

the greater the departure from $H_0$, the less likely it is that such a departure will not be detected.

# 8.1. Hypotheses and Test Procedures

- Example 8.3 (Ex. 8.1 continued)

Let us use the same experiment and test statistic *X* as previously described in the automobile bumper problem but now consider the rejection region $R_9 = \{9,10,\ldots,19,20\}$;

Since X still has a binomial distribution with parameters and p,

$$\begin{aligned}\alpha &= P(H_0 \text{ is rejected when } p = .25) \\ &= P(X \geq 9 \text{ when } X \sim \text{Bin}(20, .25)) = 1 - B(8; 20, .25) = .041\end{aligned}$$

$$\begin{aligned}\beta(.3) &= P(H_0 \text{ is not rejected when } p = .3) \\ &= P(X \leq 8 \text{ when } X \sim \text{Bin}(20, .3)) = B(8; 20, .3) = .887\end{aligned}$$

$$\beta(.5) = B(8; 20, .5) = .252$$

- Example 8.2

The drying time of a certain type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation 9 min. It is believed that drying times with an additive will remain normally distributed with $\delta = 9$. Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted.

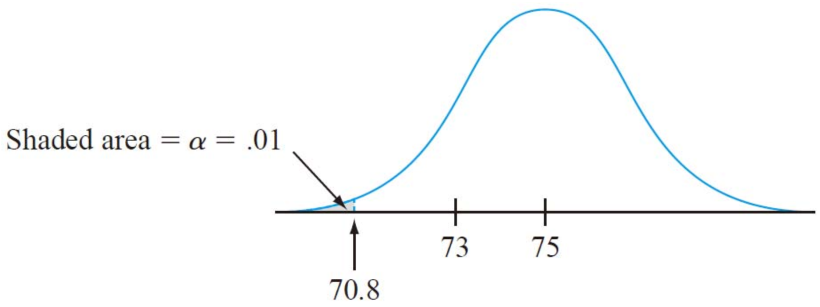
Let $\mu$ denote the true average drying time when the additive is used.

$$H_0: \mu = 75 \text{ versus } H_a: \mu < 75$$

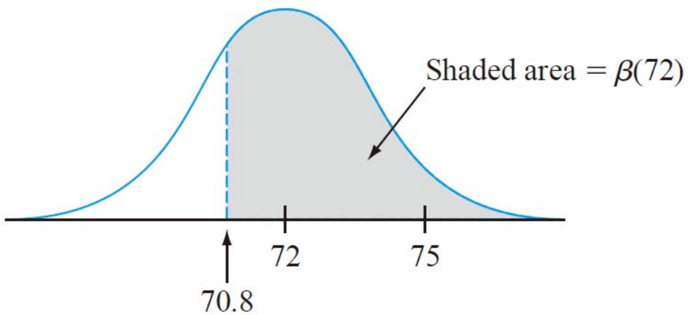
Experimental data is to consist of drying time form n = 25 test specimens. A reasonable rejection region has the form $\overline{x} \leq c$, where the cutoff value $c$ is suitably chosen, e.g., c = 70.8.

- Example 8.2 (cont')

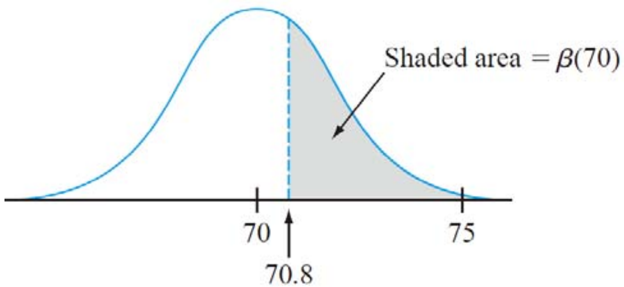
$$\begin{aligned}\alpha &= P(\text{type I error}) = P(H_0 \text{ is rejected when it is true})\\ &= P(\overline{X} \le 70.8 \text{ when } \overline{X} \sim \text{normal with } \mu_{\overline{X}} = 75,\ \sigma_{\overline{X}} = 1.8)\\ &= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = .01\end{aligned}$$




$$\begin{aligned}\beta(72) &= P(\text{type II error when } \mu = 72)\\ &= P(H_0 \text{ is not rejected when it is false because } \mu = 72)\\ &= P(\overline{X} > 70.8 \text{ when } \overline{X} \sim \text{normal with } \mu_{\overline{X}} = 72 \text{ and } \sigma_{\overline{X}} = 1.8)\\ &= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) = 1 - \Phi(-.67) = 1 - .2514 = .7486\end{aligned}$$




$$\beta(70) = 1 - \Phi\left(\frac{70.8 - 70}{1.8}\right) = .3300$$



- Example 8.4 (Ex. 8.2 continued)

The use of cutoff value c=70.8 in the paint-drying example resulted in a very small value of $\alpha$ =0.01, but rather large $\beta$'s. Consider the same experiment and test statistic with the new rejection region $\bar{x} \leq 72$

$$\begin{aligned}\alpha &= P(H_0 \text{ is rejected when it is true}) \\ &= P(\bar{X} \leq 72 \text{ when } \bar{X} \sim N(75, 1.8^2)) \\ &= \Phi\left(\frac{72-75}{1.8}\right) = \Phi(-1.67) = .0475 \approx .05\end{aligned}$$

$$\begin{aligned}\beta(72) &= P(H_0 \text{ is not rejected when } \mu = 72) \\ &= P(\bar{X} > 72 \text{ when } \bar{X} \text{ is a normal rv with mean 72 and standard deviation 1.8}) \\ &= 1 - \Phi\left(\frac{72-72}{1.8}\right) = 1 - \Phi(0) = .5\end{aligned}$$

$$\beta(70) = 1 - \Phi\left(\frac{72-70}{1.8}\right) = .1335 \qquad \beta(67) = .0027$$

- Proposition

Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then decreasing the size of the rejection region to obtain a smaller value of $\alpha$ results in a larger value of $\beta$ for any particular parameter value consistent with $H_a$.

This proposition says that once the test statistic and $n$ are fixed, there is **no rejection region** that will simultaneously make both $\alpha$ and all $\beta$'s small. A region must be chosen to effect a compromise between $\alpha$ and $\beta$.

A type I error is usually more serious than a type II error.

In the **level $\alpha$ test**, we specify the largest value of $\alpha$ (**significance level**) that can be tolerated and find a rejection region.

Typical $\alpha$ 0.10, 0.05, and 0.01

# 8.1. Hypotheses and Test Procedures

- Example 8.5

Again let μ denote the true average nicotine content of brand B cigarettes. The objective is to test $H_0$: $\mu = 1.5$ versus $H_a$: $\mu > 1.5$ based on a random sample $X_1, X_2, \ldots X_{32}$ of nicotine content. Suppose the distribution of nicotine content is known to be normal with $\delta = 0.2$.

Rather than use $\overline{X}$ itself as the test statistic, let's standardize , assuming that $H_0$ is true.

$$\text{Test statistic: } Z = \frac{\overline{X} - 1.5}{\sigma/\sqrt{n}} = \frac{\overline{X} - 1.5}{.0354}$$

$$\begin{aligned}\alpha &= P(\text{type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) \\ &= P(Z \geq c \text{ when } Z \sim N(0, 1))\end{aligned}$$

$\alpha = 0.05$, $\overline{x} \geq 1.56.$

# 8.1. Hypotheses and Test Procedures

- Homework

  Ex. 9, Ex.11, Ex.14

# 8.2. Tests About a Population Mean

- Case I: A Normal Population with Known δ

- Case II: Large-Sample Tests

- Case III: A Normal Population Distribution

The null hypothesis in all three cases will state that μ has a particular numerical value, the *null value,* which we will denote by $\mu_0$.

# 8.2. Tests About a Population Mean

- Case I: A Normal Population with Known δ

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

Test statistic value: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

| Alternative Hypothesis | Rejection Region for Level $\alpha$ Test |
| --- | --- |
| $H_a: \mu > \mu_0$ | $z \geq z_\alpha$ (upper-tailed test) |
| $H_a: \mu < \mu_0$ | $z \leq -z_\alpha$ (lower-tailed test) |
| $H_a: \mu \neq \mu_0$ | either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test) |

# 8.2. Tests About a Population Mean

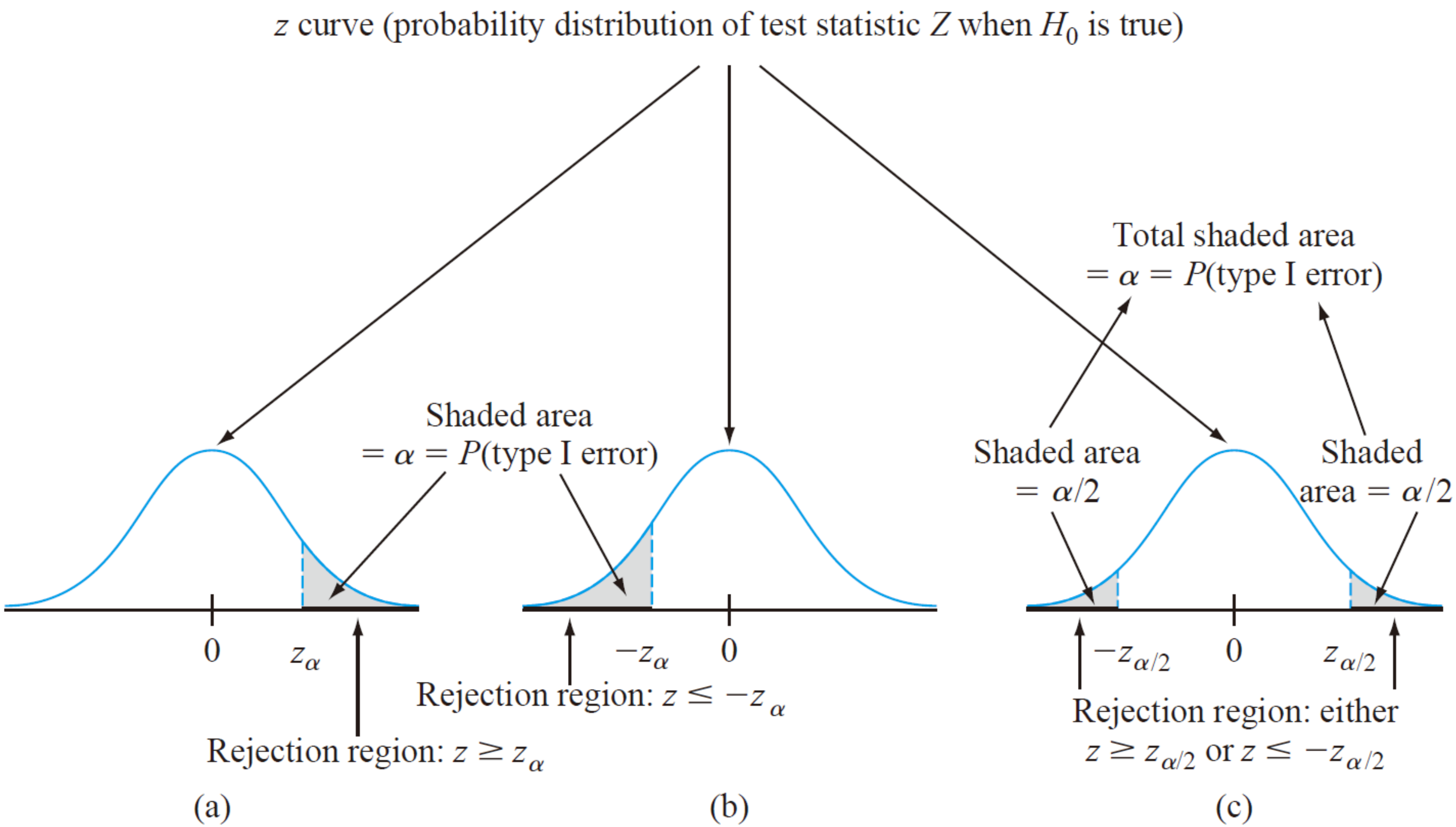



**Figure 8.2** Rejection regions for $z$ tests: (a) upper-tailed test; (b) lower-tailed test; (c) two-tailed test

# 8.2. Tests About a Population Mean

when testing hypotheses about a parameter:

1. Identify the parameter of interest and describe it in the context of the problem situation.
2. Determine the null value and state the null hypothesis. [should be done before examining the data]
3. State the appropriate alternative hypothesis. [should be done before examining the data]
4. Give the formula for the computed value of the test statistic (substituting the null value and the known values of any other parameters, but *not* those of any samplebased quantities).
5. State the rejection region for the selected significance level α.
6. Compute any necessary sample quantities, substitute into the formula for the test statistic value, and compute that value.
7. Decide whether $H_0$ should be rejected, and state this conclusion in the problem context.

- Example 8.6

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130° F. A sample of n=9 systems, when tested, yields a sample average activation temperature of 131.08° F. If the distribution of activation times is normal with standard deviation 1.5° F, does the data contradict the manufacturer's claim at significance level ?

# 8.2. Tests About a Population Mean

- Example 8.6 (Cont')

**1.** Parameter of interest: $\mu$ = true average activation temperature.

**2.** Null hypothesis: $H_0$: $\mu = 130$ (null value $= \mu_0 = 130$).

**3.** Alternative hypothesis: $H_a$: $\mu \neq 130$ (a departure from the claimed value in *either* direction is of concern).

**4.** Test statistic value:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

- Example 8.6 (Cont')

**5.** Rejection region: The form of $H_a$ implies use of a two-tailed test with rejection region *either* $z \geq z_{.005}$ *or* $z \leq -z_{.005}$. From Section 4.3 or Appendix Table A.3, $z_{.005} = 2.58$, so we reject $H_0$ if either $z \geq 2.58$ or $z \leq -2.58$.

**6.** Substituting $n = 9$ and $\overline{x} = 131.08$,

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = \frac{1.08}{.5} = 2.16$$

That is, the observed sample mean is a bit more than 2 standard deviations above what would have been expected were $H_0$ true.

**7.** The computed value $z = 2.16$ does not fall in the rejection region ($-2.58 < 2.16 < 2.58$), so $H_0$ cannot be rejected at significance level .01. The data does not give strong support to the claim that the true average differs from the design value of 130.

## 8.2. Tests About a Population Mean

- β and Sample Size Determination

Consider first the upper-tailed test with rejection region

$$z \geq z_{\alpha}, \text{ i.e., } \quad \bar{x} \geq \mu_0 + z_{\alpha} \cdot \sigma/\sqrt{n},$$

Thus $H_0$ will not rejected if $\bar{x} < \mu_0 + z_{\alpha} \cdot \sigma/\sqrt{n}$.

Denote a particular value of μ' that exceeds the null value $\mu_0$.

$$\begin{aligned}
\beta(\mu') &= P(H_0 \text{ is not rejected when } \mu = \mu') \\
&= P(\bar{X} < \mu_0 + z_{\alpha} \cdot \sigma/\sqrt{n} \text{ when } \mu = \mu') \\
&= P\left(\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} < z_{\alpha} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \text{ when } \mu = \mu'\right) \\
&= \Phi\left(z_{\alpha} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

# 8.2. Tests About a Population Mean

- More generally,

consider the two restrictions $P(\text{type I error}) = \alpha$ and $\beta(\mu') = \beta$ for specified $\alpha$, $\mu'$, and $\beta$. Then for an upper-tailed test, the sample size $n$ should be chosen to satisfy

$$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \beta$$

This implies that

$$-z_\beta = \begin{array}{l} z \text{ critical value that} \\ \text{captures lower-tail area } \beta \end{array} = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}$$

# 8.2. Tests About a Population Mean

| Alternative Hypothesis | Type II Error Probability $\beta(\mu')$ for a Level $\alpha$ Test |
| --- | --- |
| $H_a$: $\mu > \mu_0$ | $\Phi\left(z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$ |
| $H_a$: $\mu < \mu_0$ | $1 - \Phi\left(-z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$ |
| $H_a$: $\mu \neq \mu_0$ | $\Phi\left(z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$ |

where $\Phi(z) =$ the standard normal cdf.

The sample size $n$ for which a level $\alpha$ test also has $\beta(\mu') = \beta$ at the alternative value $\mu'$ is

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a one-tailed (upper or lower) test} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a two-tailed test (an approximate solution)} \end{cases}$$

# 8.2. Tests About a Population Mean

- Example 8.7

Let $\mu$ denote the true average tread life of a certain type of tire. Consider testing $H_0$: $\mu = 30{,}000$ versus $H_a$: $\mu > 30{,}000$ based on a sample of size $n = 16$ from a normal population distribution with $\sigma = 1500$. A test with $\alpha = .01$ requires $z_\alpha = z_{.01} = 2.33$. The probability of making a type II error when $\mu = 31{,}000$ is

$$\beta(31{,}000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Since $z_{.1} = 1.28$, the requirement that the level .01 test also have $\beta(31{,}000) = .1$ necessitates

$$n = \left[\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}\right]^2 = (-5.42)^2 = 29.32$$

n=30

# 8.2. Tests About a Population Mean

- Case II: Large-Sample Tests

When the sample size n is large (n>40), the $z$ tests for case I are easily modified to yield valid test procedures without requiring either a normal population distribution or known σ.

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

has *approximately* a standard normal distribution

Similar to the Case I.

- Case III: A Normal Population Distribution

When $n$ is small, the Central Limit Theorem (CLT) can no longer be invoked to justify the use of a large-sample test.

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a $t$ distribution with degrees of freedom (df).

# 8.2. Tests About a Population Mean

**The One-Sample $t$ Test**

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

Test statistic value: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

| Alternative Hypothesis | Rejection Region for a Level $\alpha$ Test |
|---|---|
| $H_a$: $\mu > \mu_0$ | $t \geq t_{\alpha,n-1}$ (upper-tailed) |
| $H_a$: $\mu < \mu_0$ | $t \leq -t_{\alpha,n-1}$ (lower-tailed) |
| $H_a$: $\mu \neq \mu_0$ | either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$ (two-tailed) |

Compared to the table in page 328.

# 8.2. Tests About a Population Mean

- Homework

  Ex. 15, Ex. 17, Ex. 18