

8

Tests of Hypotheses Based on a Single Sample

Copyright © Cengage Learning. All rights reserved.

8.1

Hypotheses and Test Procedures

Copyright © Cengage Learning. All rights reserved.

Introduction

- **Parameter** can be **estimated** from **sample data** either by
 - single number (**Point Estimate**) or
 - entire interval of plausible values (**Confidence Interval**)
- However, **objective of investigation** is **not to estimate parameter** but **to decide which of two contradictory claims** about **parameter** is correct
- **Methods** for accomplishing this comprise the **part of statistical inference** called

การทดสอบสมมติฐาน (Hypothesis Testing)

3

Hypotheses and Test Procedures

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

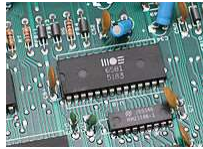
4

Hypotheses and Test Procedures

- One example of a hypothesis is claim $\mu = 0.75$, where μ is the true average inside diameter of a certain type of PVC pipe.



- Another example is statement $p < 0.10$, where p is proportion of defective circuit boards among all circuit boards produced by a certain manufacturer.



5

Hypotheses and Test Procedures

- In any hypothesis-testing problem, there are two contradictory hypotheses under consideration.
- One hypothesis might be
 - the claim $\mu = 0.75$ and
 - the other $\mu \neq 0.75$, or
- the two contradictory statements might be
 - $p \geq 0.10$ and
 - $p < 0.10$.



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

7

Hypotheses and Test Procedures

การพิจารณาคดีอาญา

- There is familiar analogy to this in **criminal trial**.
- **One claim** is **assertion** that **accused individual** is **innocent**.
- In the **U.S. judicial system**, this is **claim** that is **initially believed** to be **true**.
- Only in the face of **strong evidence** to **contrary** should **jury reject this claim** in favor of **alternative assertion** that **accused** is **guilty**.



- In this sense, **claim of innocence** is **avored** or **protected hypothesis**, and **burden of proof** is placed on those who **believe in** the **alternative claim**.

8

Hypotheses and Test Procedures

- Similarly, **in testing statistical hypotheses**, problem will be formulated so that one of **claims** is **initially favored**.
- This **initially favored claim** will **not be rejected** in favor of **alternative claim** unless **sample evidence contradicts it** and provides **strong support** for **alternative assertion**.

9

Hypotheses and Test Procedures

Definition

- **Null Hypothesis**, denoted by H_0 , is **claim** that is **initially assumed** to be **true** (“prior belief” claim).
- **Alternative Hypothesis**, denoted by H_a , is **assertion** that is **contradictory** to H_0 .
- **Null hypothesis** will **be rejected** in favor of the **alternative hypothesis** only if **sample evidence** suggests that **H_0 is false**.
- If **sample** does not strongly contradict H_0 , we will continue to believe in plausibility of **null hypothesis**.
- **Two possible conclusions** from hypothesis-testing analysis are then **reject H_0** or **fail to reject H_0** .

10


Hypotheses and Test Procedures

- **Test of hypotheses** is **method** for **using sample data** to **decide** whether **null hypothesis** should **be rejected**.
- Thus we might test
 $H_0: \mu = 0.75$ against the alternative
 $H_a: \mu \neq 0.75$.
- Only if **sample data** **strongly suggests** that μ is something other than 0.75 should null hypothesis be rejected.
- In **absence** of such evidence, H_0 should **not be rejected**, since it is still **quite plausible**.



11

Hypotheses and Test Procedures

- Sometimes investigator does not want to accept particular assertion unless and until data can provide strong support for assertion.
 - As an example, suppose company is considering putting new type of coating on bearings that it produces.
- 
- True average wear life with current coating is known to be 1000 hours.
 - With μ denoting true average life for new coating, company would not want to make a change unless evidence strongly suggested that μ exceeds 1000.

12

Hypotheses and Test Procedures

- Appropriate problem formulation would involve testing $H_0: \mu = 1000$ against $H_a: \mu > 1000$.
- Conclusion that change is justified is identified with H_a , and it would take conclusive evidence to justify rejecting H_0 and switching to the new coating.
- Scientific research often involves trying to decide whether current theory should be replaced by more plausible and satisfactory explanation of phenomenon under investigation.

13

Hypotheses and Test Procedures

- Conservative approach is to identify current theory with H_0 and researcher's alternative explanation with H_a .
- Rejection of current theory will then occur only when evidence is much more consistent with new theory.
- In many situations, H_a is referred to as “researcher's hypothesis,” since it is claim that the researcher would really like to validate.

14

Hypotheses and Test Procedures

- The word **null** means “of no value, effect, or consequence,” which suggests that H_0 should be identified with the hypothesis of no change (from current opinion), no difference, no improvement, and so on.



- Suppose, for example, that 10% of all circuit boards produced by a certain manufacturer during a recent period were defective.
- Engineer has suggested a change in the production process in the belief that it will result in reduced defective rate.

15

Hypotheses and Test Procedures

- Let p denote true proportion of defective boards resulting from changed process.
- Then research hypothesis, on which burden of proof is placed, is assertion that $p < 0.10$.
- Thus alternative hypothesis is $H_a: p < 0.10$.
- In our treatment of hypothesis testing, H_0 will generally be stated as equality claim.
- If θ denotes parameter of interest, the null hypothesis will have the form $H_0: \theta = \theta_0$, where θ_0 is specified number called null value of parameter (value claimed for θ by null hypothesis).

16

Hypotheses and Test Procedures



- As example, consider circuit board situation just discussed.
- Suggested alternative hypothesis was $H_a: p < 0.10$, claim that defective rate is reduced by process modification.
- Natural choice of H_0 in this situation is claim that $p \geq 0.10$, according to which new process is either no better or worse than one currently used.
- We will instead consider $H_0: p = 0.10$ versus $H_a: p < 0.10$.

17

Hypotheses and Test Procedures

- Rationale for using this simplified null hypothesis is that any reasonable decision procedure for deciding between

$$H_0: p = 0.10 \text{ and } H_a: p < 0.10$$

will also be reasonable for deciding between the claim that $p \geq 0.10$ and H_a .

- The use of simplified H_0 is preferred because it has certain technical benefits, which will be apparent shortly.

18

Hypotheses and Test Procedures

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

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

19

Hypotheses and Test Procedures



- For example, let σ denote standard deviation of distribution of inside diameters (inches) for certain type of metal sleeve.
- If decision was made to use sleeve unless sample evidence conclusively demonstrated that $\sigma > 0.001$, the appropriate hypotheses would be

$$H_0: \sigma = 0.001. \text{ versus } H_a: \sigma > 0.001.$$

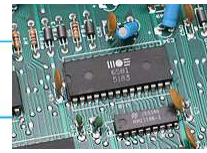
- The number θ_0 that appears in both H_0 and H_a (separates the alternative from the null) is called the **null value**.

20

Test Procedures

21

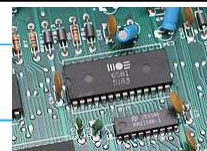
Test Procedures



- **Test procedure** is a **rule**, based on **sample data**, for deciding **whether to reject H_0** .
- Test of $H_0: p = 0.10$ versus $H_a: p < 0.10$ in **circuit board problem** might be based on examining **random sample of $n = 200$ boards**.
- Let X denote **number of defective boards in the sample**, a **binomial random variable**; x represents **observed value of X** .

22

Test Procedures



- If H_0 is true, $E(X) = np = 200(0.10) = 20$, whereas we can expect **fewer than 20 defective boards if H_a is true**.
- 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**.

23

Test Procedures



This procedure has **two** constituents:

- (1) **Test Statistic**, or function of sample data used to **make a decision**, and
- (2) **Rejection Region** consisting of those x values for which H_0 will be rejected in favor of H_a .
 - For rule just suggested, the **rejection region** consists of $x = 0, 1, 2, \dots$, and 15.
 - H_0 will **not be rejected** if $x = 16, 17, \dots, 199$, or 200.

24

Test Procedures

A test procedure is specified by the following:

1. **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. **Rejection Region**, the set of all test statistic values for which H_0 will be rejected
 - **Null hypothesis** will then be rejected if and only if the observed or computed **test statistic value falls in the rejection region**.

25

Example :



- Suppose cigarette manufacture claims that average nicotine content μ of brand B cigarettes in (at most) 1.5 mg.
- It would be unwise to reject manufacturer's claim without strong contradictory evidence, so an appropriate problem formulation is to test

$$H_0 : \mu = 1.5 \text{ versus } H_a : \mu > 1.5.$$

- Consider decision rule based on analyzing a random sample of 32 cigarettes.
- Let \bar{X} denote the sample average nicotine content.
- If H_0 is true, $E(\bar{X}) = \mu = 1.5$ whereas
If H_0 is false, we expect \bar{X} to exceed 1.5
- Thus we might use \bar{X} as a test statistic along with rejection region $\bar{x} \geq 1.6$

26

Errors in Hypothesis Testing

27

Errors in Hypothesis Testing

- The basis for choosing particular rejection region lies in consideration of errors that one might be faced with in drawing conclusion.
- Consider **rejection region $x \leq 15$** in circuit board problem.

$$E(X) = np = 200(0.10) = 20$$
- Even when $H_0: p = 0.10$ is true, it might happen that unusual sample results in $x = 13$, so that H_0 is erroneously rejected.

$$E(X) = np = 200(0.10) = 20$$
- On the other hand, even when $H_a: p < 0.10$ is true, unusual sample might yield $x = 20$, in which case H_0 would not be rejected—again an incorrect conclusion.

28

Errors in Hypothesis Testing

- Thus it is possible that
 H_0 may be rejected when it is true
 or that
 H_0 may not be rejected when it is false.
- These possible errors are not consequences of a foolishly chosen rejection region.

29

Errors in Hypothesis Testing

Definition

- **Type I error** consists of **rejecting** null hypothesis H_0 when H_0 is true.
- **Type II error** involves **not rejecting** H_0 when H_0 is false.

31

Errors in Hypothesis Testing

$$H_0 : \mu = 1.5, \quad H_a : \mu > 1.5$$

- Thus when $\mu = 1.5$ in the nicotine situation, \bar{x} may be larger than 1.5, resulting in **erroneous rejection** of H_0 .

$$H_0 : \mu = 1.6, \quad H_a : \mu > 1.6$$

- Alternatively, it may be that $\mu = 1.6$ yet \bar{x} much smaller than this is observed, leading to **type II error**.

Type I error consists of **rejecting** null hypothesis H_0 when H_0 is true.
Type II error involves **not rejecting** H_0 when H_0 is false.

34

Errors in Hypothesis Testing

- Instead of demanding error-free procedures, we must seek procedures for which either type of error is unlikely to occur.
- That is, good procedure is one for which probability of making either type of error is small.
- The choice of particular rejection region cutoff value fixes probabilities of type I and type II errors.

35


Errors in Hypothesis Testing

- These error probabilities are traditionally denoted by α and β , respectively.
- Because H_0 specifies a unique value of parameter, there is a single value of α .
- However, there is a different value of β for each value of parameter consistent with H_a .

36

Example



- A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph (16.09 km/h) 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.

37

Example



cont'd

- Intuitively, H_0 should be rejected if substantial number of crashes show no damage.
- Consider the following test procedure:

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

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

38

Example

cont'd

- This **rejection region** is called *upper-tailed* because it consists only of **large values** of test statistic.
- When H_0 is true, X has **binomial probability distribution** with $n = 20$ and $p = 0.25$. Then

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

$$= P(X \geq 8 \text{ when } X \sim \text{Bin}(20, 0.25)) = 1 - B(7; 20, 0.25)$$

$$= 1 - 0.898$$

$$= 0.102$$

↓
10.2%

	p																n = 20
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99		
0	.818	.358	.122	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
1	.983	.736	.392	.069	.024	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	
2	.999	.925	.677	.206	.091	.035	.004	.000	.000	.000	.000	.000	.000	.000	.000	.000	
3	1.000	.984	.867	.411	.225	.107	.016	.001	.000	.000	.000	.000	.000	.000	.000	.000	
4	1.000	.997	.957	.630	.415	.238	.051	.006	.000	.000	.000	.000	.000	.000	.000	.000	
5	1.000	1.000	.989	.804	.617	.416	.126	.021	.002	.000	.000	.000	.000	.000	.000	.000	
6	1.000	1.000	.998	.913	.786	.608	.250	.058	.006	.000	.000	.000	.000	.000	.000	.000	
7	1.000	1.000	1.000	.968	.898	.772	.416	.132	.021	.001	.000	.000	.000	.000	.000	.000	
8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000	.000	

Example

cont'd

- 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.



Type I error

Example

cont'd

- In contrast to α , there is not single β .
- Instead, there is different β for each different p that exceeds 0.25.
- Thus there is value of β for $p = 0.3$
(in which case $X \sim \text{Bin}(20, 0.3)$), another value of β for $p = 0.5$, and so on.

41

Example

cont'd

For example,

$$\begin{aligned}
 \beta(0.3) &= P(\text{type II error when } p = 0.3) \\
 &= P(H_0 \text{ is not rejected when } H_0 \text{ is false because } p = 0.3) \\
 &= P(X \leq 7 \text{ when } X \sim \text{Bin}(20, 0.3)) = B(7; 20, 0.3) = 0.772
 \end{aligned}$$

		p														
		0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
n = 20	0	.818	.358	.122	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.983	.736	.392	.069	.024	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.999	.925	.677	.206	.091	.035	.004	.000	.000	.000	.000	.000	.000	.000	.000
	3	1.000	.984	.867	.411	.225	.107	.016	.001	.000	.000	.000	.000	.000	.000	.000
	4	1.000	.997	.957	.630	.415	.238	.051	.006	.000	.000	.000	.000	.000	.000	.000
	5	1.000	1.000	.989	.804	.617	.416	.126	.021	.002	.000	.000	.000	.000	.000	.000
	6	1.000	1.000	.998	.913	.786	.608	.250	.058	.006	.000	.000	.000	.000	.000	.000
	7	1.000	1.000	1.000	.968	.898	.772	.416	.132	.021	.001	.000	.000	.000	.000	.000
8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000	

n = 20

- When p is actually 0.3 rather than 0.25 (a “small” departure from H_0), roughly 77% of all experiments of this type would result in H_0 being incorrectly not rejected!

42

Example

cont'd

- The accompanying table displays β for selected values of p (each calculated for rejection region R_g).

p	0.3	0.4	0.5	0.6	0.7	0.8
$\beta(p)$	0.772	0.416	0.132	0.021	0.001	0.000

- Clearly, β decreases as value of p moves farther to the right of the null value 0.25.
- Intuitively, the greater the departure from H_0 , the less likely it is that such departure will not be detected.

43

Example ...

cont'd

- Let us use the same experiment and test statistic X as previously described in automobile bumper problem but now consider

rejection region $R_g = \{9, 10, \dots, 20\}$

- Since X still has a binomial distribution with parameter $n=20$ and p .

46

Example ...

cont'd

Type I error

$$\begin{aligned}\alpha &= P(\text{H}_0 \text{ is rejected when } p = 0.25) \\ &= P(X \geq 9 \text{ when } X \sim \text{Bin}(20, 0.25)) = 1 - B(8; 20, 0.25) = 0.041\end{aligned}$$

$n = 20$		0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
X	8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000
	9	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

- Type I error probability has been decreased by using new rejection region.

47

Example ...

cont'd

- However, a price has been paid for this decrease :

$$\begin{aligned}\beta(0.3) &= P(\text{H}_0 \text{ is not rejected when } p = 0.3) \\ &= P(X \leq 8 \text{ when } X \sim \text{Bin}(20, 0.3)) = B(8; 20, 0.3) = 0.887\end{aligned}$$

$n = 20$		0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
X	8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000
	9	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

$$\begin{aligned}\beta(0.5) &= P(\text{H}_0 \text{ is not rejected when } p = 0.5) \\ &= P(X \leq 8 \text{ when } X \sim \text{Bin}(20, 0.5)) = B(8; 20, 0.5) = 0.252\end{aligned}$$

48

Example ...

cont'd

	p	0.3	0.5
$\alpha = 0.102$	$R8 \rightarrow \beta(p)$	0.772	0.132
$\alpha = 0.041$	$R9 \rightarrow \beta(p)$	0.887	0.252

- Both these β s are larger than corresponding error probabilities 0.772 and 0.132 for region R_8
- This is not surprising; α is computed by summing over probabilities of test statistic values in rejection region, where as β is probability that X falls in complement of rejection region.
- Making rejection region smaller must therefore decrease α while increasing β for any fixed alternative value of parameter.

49

Errors in Hypothesis Testing

Proposition

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

50

Errors in Hypothesis Testing

- This proposition says that once test statistic and n are fixed, there is no rejection region that will simultaneously make both α and all β 's small.
- Region must be chosen to effect a compromise between α and β .
- Because of suggested guidelines for specifying H_0 and H_a , a type I error is usually more serious than a type II error (this can always be achieved by proper choice of hypotheses).

51

Errors in Hypothesis Testing

- Approach adhered to by most statistical practitioners is then to specify the largest value of that can be tolerated and find rejection region having that value of α rather than anything smaller.



This makes β as small as possible subject to the bound on α .



Resulting value of α is often referred to as **significance level** of test.

- Traditional levels of significance are 0.10, 0.05, and 0.01, though level in any particular problem will depend on the seriousness of type I error.



The more serious this error, the smaller should be significance level.

Errors in Hypothesis Testing

- The corresponding test procedure is called a **level α test** e.g.,
 - level 0.05 test or
 - level 0.01 test.
- A test with significance level α is one for which the type I error probability is controlled at specified level.

53

Example

- Again let μ denote true average nicotine content of brand B cigarettes.
- The objective is to test

$$H_0: \mu = 1.5 \text{ versus } H_a: \mu > 1.5$$
 based on a random sample X_1, X_2, \dots, X_{32} of nicotine content.
- Suppose distribution of nicotine content is known to be normal with $\sigma = 0.20$.
- Then \bar{X} is normally distributed with

mean value $\mu_{\bar{X}} = \mu$ and
 standard deviation $\sigma_{\bar{X}} = \frac{0.20}{\sqrt{32}} = 0.0354$

54

Distribution of Sample Mean (from Chapter 5)

Proposition

- Let X_1, X_2, \dots, X_n be a random sample from a distribution with mean value μ and standard deviation σ . Then

1. $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. $V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

55

Example

$$Z = \frac{X - \mu}{\sigma} \Rightarrow Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ cont'd}$$

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

$$\text{Test statistic : } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow Z = \frac{\bar{X} - 1.5}{\frac{0.20}{\sqrt{32}}} = \frac{\bar{X} - 1.5}{0.0354}$$

- Z expresses the distance between \bar{X} and its expected value (μ) when H_0 is true as some number of standard deviations.
- For example, $z = 3$ results from \bar{x} that is 3 standard deviations larger than we would have expected it to be were H_0 true.

56

Example

cont'd

That is, the form of the rejection region is $z \geq c$.

Let's now determine c so that $\alpha = 0.05$.

When H_0 is true, Z has a standard normal distribution. Thus

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

- Value c must capture upper-tail area 0.05 under the z curve.
- So, directly from Appendix Table A.3,

		$\Phi(z) = P(Z \leq z)$								
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$$C = z_{0.05} = 1.645.$$

58

Example

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ cont'd}$$

- Notice that

$$Z = \frac{\bar{X} - 1.5}{0.0354}$$

$$z \geq 1.645 \quad \text{is equivalent to} \quad \bar{x} - 1.5 \geq (0.0354)(1.645)$$

$$\bar{x} \geq 1.56$$

- Then β involves the probability that $\bar{X} < 1.56$ and can be calculated for any μ greater than 1.5.

59



End of Section 8.1

60