ACE 261
Fall 2002
Prof. Katchova

Lecture 9

Hypothesis Testing

Outline: Hypothesis Testing

- Developing Null and Alternative Hypotheses
- Type I and Type II Errors
- One-Tailed Tests About a Population Mean: Large-Sample Case
- Two-Tailed Tests About a Population Mean: Large-Sample Case
- Tests About a Population Mean: Small-Sample Case
- Tests About a Population Proportion
- Hypothesis Testing and Decision Making
- Calculating the Probability of Type II Errors
- Determining the Sample Size for a Hypothesis Test about a Population Mean

2

Developing Null and Alternative Hypotheses

- Hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.
- The null hypothesis, denoted by H_0, is a tentative assumption about a population parameter.
- The alternative hypothesis, denoted by H_a, is the opposite of what is stated in the null hypothesis.
- Hypothesis testing is similar to a criminal trial. The hypotheses are:

  H_0: The defendant is innocent

  H_a: The defendant is guilty

3

Developing Null and Alternative Hypotheses

- Testing Research Hypotheses
  - The research hypothesis should be expressed as the alternative hypothesis.
  - The conclusion that the research hypothesis is true comes from sample data that contradict the null hypothesis.

4

Developing Null and Alternative Hypotheses

- Testing the Validity of a Claim
  - Manufacturers' claims are usually given the benefit of the doubt and stated as the null hypothesis.
  - The conclusion that the claim is false comes from sample data that contradict the null hypothesis.

5

Developing Null and Alternative Hypotheses

- Testing in Decision-Making Situations
  - A decision maker might have to choose between two courses of action, one associated with the null hypothesis and another associated with the alternative hypothesis.
  - Example: Accepting a shipment of goods from a supplier or returning the shipment of goods to the supplier.

6

A Summary of Forms for Null and Alternative Hypotheses

- The equality part of the hypotheses always appears in the null hypothesis.
- In general, a hypothesis test about the value of a population mean μ must take one of the following three forms (where μ_0 is the hypothesized value of the population mean).

H_0: $\mu \geq \mu_0$	H_0: $\mu \leq \mu_0$	H_0: $\mu = \mu_0$
H_a: $\mu < \mu_0$	H_a: $\mu > \mu_0$	H_a: $\mu \neq \mu_0$

7

Example

- A major west coast city provides one of the most comprehensive emergency medical services in the world. Operating in a multiple hospital system with approximately 20 mobile medical units, the service goal is to respond to medical emergencies with a mean time of 12 minutes or less.
- The director of medical services wants to formulate a hypothesis test that could use a sample of emergency response times to determine whether or not the service goal of 12 minutes or less is being achieved.

8

Example

Hypotheses	Conclusion and Action
H_0: $\mu \leq 12$	The emergency service is meeting the response goal; no follow-up action is necessary.
H_a: $\mu > 12$	The emergency service is not meeting the response goal; appropriate follow-up action is necessary.

Where: μ = mean response time for the population of medical emergency requests.

9

Type I and Type II Errors

- Since hypothesis tests are based on sample data, we must allow for the possibility of errors.
- A Type I error is rejecting H_0 when it is true.
- A Type II error is accepting H_0 when it is false.
- The person conducting the hypothesis test specifies the maximum allowable probability of making a Type I error, denoted by α and called the level of significance.
- Generally, we cannot control for the probability of making a Type II error, denoted by β.
- Statistician avoids the risk of making a Type II error by using "do not reject H_0" and not "accept H_0".

10

Type I and Type II Errors

Conclusion	Population Condition: H_0 True ($\mu \leq 12$)	Population Condition: H_a True ($\mu > 12$)
Accept H_0 (Conclude $\mu \leq 12$)	Correct Conclusion	Type II Error
Reject H_0 (Conclude $\mu > 12$)	Type I Error	Correct Conclusion

11

The Use of p-Values

- The *p*-value is the probability of obtaining a sample result that is at least as unlikely as what is observed.
- The *p*-value can be used to make the decision in a hypothesis test by noting that:
  - if the *p*-value is less than the level of significance α, the value of the test statistic is in the rejection region.
  - if the *p*-value is greater than or equal to α, the value of the test statistic is not in the rejection region.
- Reject H_0 if the *p*-value < α.

12

The Steps of Hypothesis Testing

- Determine the appropriate hypotheses.
- Select the test statistic for deciding whether or not to reject the null hypothesis.
- Specify the level of significance α for the test.
- Use α to develop the rule for rejecting H_0.
- Collect the sample data and compute the value of the test statistic.
  a) Compare the test statistic to the critical value(s) in the rejection rule, or
  b) Compute the *p*-value based on the test statistic and compare it to α to determine whether or not to reject H_0.

13

One-Tailed Tests about a Population Mean: Large-Sample Case (n $\geq$ 30)

Hypotheses

H_0: $\mu \leq \mu_0$ or H_0: $\mu \geq \mu_0$

H_a: $\mu > \mu_0$ $\quad$ H_a: $\mu < \mu_0$

Test Statistic

<u>σ Known</u> <u>σ Unknown</u>

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

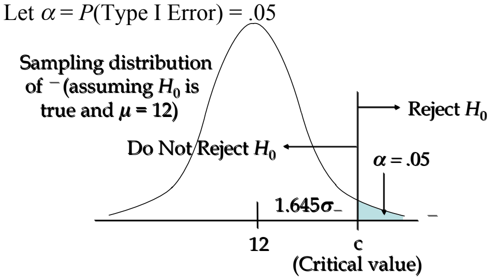
Rejection Rule

Reject H_0 if $z > z_\alpha$ $\quad$ Reject H_0 if $z < -z_\alpha$

14

One-Tailed Test about a Population Mean: Large *n*

Let $\alpha = P$(Type I Error) = .05



15

One-Tailed Test about a Population Mean: Large *n*

- Let $n = 40$, xbar = 13.25 minutes, $s = 3.2$ minutes
  (The sample standard deviation *s* can be used to estimate the population standard deviation σ.)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{13.25 - 12}{3.2 / \sqrt{40}} = 2.47$$

Since 2.47 > 1.645, we reject H_0.

<u>Conclusion</u>**:** We are 95% confident that Metro EMS is not meeting the response goal of 12 minutes; appropriate action should be taken to improve service.

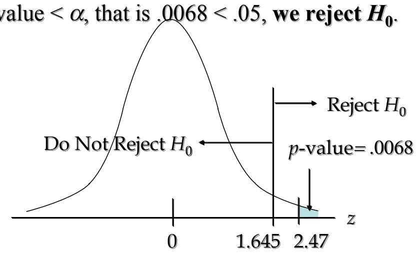
16

Using the p-value to Test Hypothesis

Recall that $z = 2.47$ for xbar = 13.25.

Then *p*-value = .0068.

Since *p*-value < α, that is .0068 < .05, **we reject H_0**.



17

Two-Tailed Tests about a Population Mean: Large-Sample Case

- Hypothesis

  H_0: $\mu = \mu_0$

  H_a: $\mu \neq \mu_0$

- Test statistic <u>σ Known</u> <u>σ Unknown</u>

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

- Rejection rule

  Reject H_0 if $|z| > z_{\alpha/2}$

18

One-Tail Test about a Population Mean: Large n

- The production line for Glow toothpaste is designed to fill tubes of toothpaste with a mean weight of 6 ounces.
- Periodically, a sample of 30 tubes will be selected in order to check the filling process. Quality assurance procedures call for the continuation of the filling process if the sample results are consistent with the assumption that the mean filling weight for the population of toothpaste tubes is 6 ounces; otherwise the filling process will be stopped and adjusted.

Two-Tail Test about a Population Mean: Large n

- A hypothesis test about the population mean can be used to help determine when the filling process should continue operating and when it should be stopped and corrected.
  - Hypothesis

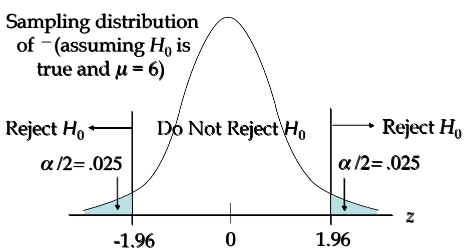
$$H_0: \ \mu = 6$$
$$H_a: \ \mu \neq 6$$

  - Rejection rule

Assuming a .05 level of significance,

Reject H_0 if $z < -1.96$ or if $z > 1.96$

Two-Tailed Test about a Population Mean: Large n



One-Tail Test about a Population Mean: Large n

- **Assume that a sample of 30 toothpaste tubes provides a sample mean of 6.1 ounces and standard deviation of 0.2 ounces.**

Let $n = 30$, xbar = 6.1 ounces, $s = .2$ ounces

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{6.1 - 6}{.2/\sqrt{30}} = 2.74$$

Since 2.74 > 1.96, we reject H_0.

Conclusion: We are 95% confident that the mean filling weight of the toothpaste tubes is not 6 ounces. The filling process should be stopped and the filling mechanism adjusted.

Using the p-Value for a Two-Tailed Hypothesis Test

- Suppose we define the *p*-value for a two-tailed test as **double** the area found in the tail of the distribution.

With $z = 2.74$, the standard normal probability table shows there is a .5000 - .4969 = .0031 probability of a difference larger than .1 in the upper tail of the distribution.

Considering the same probability of a larger difference in the lower tail of the distribution, we have

p-value = 2(.0031) = .0062

The p-value .0062 is less than $\alpha = .05$, so H_0 is rejected.

Confidence Interval Approach to a Two-Tailed Test

- Select a simple random sample from the population and use the value of the sample mean to develop the confidence interval for the population mean μ.
- If the confidence interval contains the hypothesized value μ_0, do not reject H_0. Otherwise, reject H_0.

Confidence Interval Approach to a Two-Tailed Test

The 95% confidence interval for μ is

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6.1 \pm 1.96(.2/\sqrt{30}) = 6.1 \pm .0716$$

or 6.0284 to 6.1716

Since the hypothesized value for the population mean, $\mu_0 = 6$, is not in this interval, the hypothesis-testing conclusion is that the null hypothesis, H_0: $\mu = 6$, can be rejected.

25

Tests about a Population Mean: Small-Sample Case (n < 30)

- Test statistics

	σ Known	σ Unknown
	$t = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

This test statistic has a t distribution with n - 1 degrees of freedom.

- Rejection rule

	One-Tailed	Two-Tailed
H_0: $\mu < \mu_0$	Reject H_0 if $t > t_\alpha$	
H_0: $\mu > \mu_0$	Reject H_0 if $t < -t_\alpha$	
H_0: $\mu = \mu_0$		Reject H_0 if $\lvert t\rvert > t_{\alpha/2}$

26

p -Values and the t Distribution

- The format of the t distribution table provided in most statistics textbooks does not have sufficient detail to determine the exact p-value for a hypothesis test.
- However, we can still use the t distribution table to identify a range for the p-value.
- An advantage of computer software packages is that the computer output will provide the p-value for the t distribution.

27

One-Tailed Test about a Population Mean: Small n

- A State Highway Patrol periodically samples vehicle speeds at various locations on a particular roadway. The sample of vehicle speeds is used to test the hypothesis

$$H_0:\ \mu \leq 65.$$

The locations where H_0 is rejected are deemed the best locations for radar traps.

At Location F, a sample of 16 vehicles shows a mean speed of 68.2 mph with a standard deviation of 3.8 mph. Use an α = .05 to test the hypothesis.

28

One-Tail Test about a Population Mean: Small n

Let n = 16, xbar = 68.2 mph, s = 3.8 mph

α = .05, d.f. = 16-1 = 15, t_α = 1.753

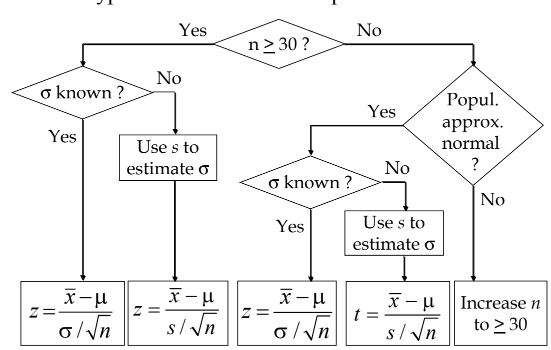
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{68.2 - 65}{3.8/\sqrt{16}} = 3.37$$

Since 3.37 > 1.753, we reject H_0.

Conclusion: We are 95% confident that the mean speed of vehicles at Location F is greater than 65 mph. Location F is a good candidate for a radar trap.

29

Summary of Test Statistics to be Used in a Hypothesis Test about a Population Mean



30

A Summary of Forms for Null and Alternative Hypotheses about a Population Proportion

- The equality part of the hypotheses always appears in the null hypothesis.
- In general, a hypothesis test about the value of a population proportion *p* must take one of the following three forms (where p_0 is the hypothesized value of the population proportion).

$H_0\!: p \geq p_0$ $\quad H_0\!: p \leq p_0 \quad H_0\!: p = p_0$

$H_a\!: p < p_0$ $\quad H_a\!: p > p_0 \quad H_a\!: p \neq p_0$

31

Tests about a Population Proportion: Large-Sample Case

- Test Statistic $z = \dfrac{\bar{p} - p_0}{\sigma_{\bar{p}}}$

  where: $\sigma_{\bar{p}} = \sqrt{\dfrac{p_0(1-p_0)}{n}}$

- Rejection rule

	One-Tailed	Two-Tailed
$H_0\!: p \leq p_0$	Reject H_0 if $z > z_\alpha$	
$H_0\!: p \geq p_0$	Reject H_0 if $z < -z_\alpha$	
$H_0\!: p = p_0$		Reject H_0 if $\lvert z \rvert > z_{\alpha/2}$

32

Two-Tailed Test about a Population Proportion: Large *n*

- For a Christmas and New Year's week, the National Safety Council estimated that 500 people would be killed and 25,000 injured on the nation's roads. The NSC claimed that 50% of the accidents would be caused by drunk driving.
- A sample of 120 accidents showed that 67 were caused by drunk driving. Use these data to test the NSC's claim with $\alpha = 0.05$.

33

Two-Tailed Test about a Population Proportion: Large *n*

- Hypothesis
  - $H_0\!: p = .5$
  - $H_a\!: p \neq .5$
- Test Statistic

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.5(1-.5)}{120}} = .045644$$

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{(67/120) - .5}{.045644} = 1.278$$

34

Two-Tailed Test about a Population Proportion: Large *n*

- Rejection rule
  - Reject H_0 if $z < -1.96$ or $z > 1.96$
- Conclusion
  - Do not reject H_0.
    - For $z = 1.278$, the *p*-value is .201. If we reject H_0, we exceed the maximum allowed risk of committing a Type I error (*p*-value > .050).

35

Hypothesis Testing and Decision Making

- In many decision-making situations the decision maker may want, and in some cases may be forced, to take action with both the conclusion do not reject H_0 and the conclusion reject H_0.
- In such situations, it is recommended that the hypothesis-testing procedure be extended to include consideration of making a Type II error.

36

Calculating the Probability of a Type II Error

1. Formulate the null and alternative hypotheses.
2. Use the level of significance α to establish a rejection rule based on the test statistic.
3. Using the rejection rule, solve for the value of the sample mean that identifies the rejection region.
4. Use the results from step 3 to state the values of the sample mean that lead to the acceptance of H_0; this defines the acceptance region.
5. Using the sampling distribution of xbar for any value of μ from the alternative hypothesis, and the acceptance region from step 4, compute the probability that the sample mean will be in the acceptance region.

37

Calculating the Probability of a Type II Error

1. Hypotheses are: H_0: $\mu \leq 12$ and H_a: $\mu > 12$
2. Rejection rule is: Reject H_0 if $z > 1.645$
3. Value of the sample mean that identifies the rejection region:

$$z = \frac{\bar{x} - 12}{3.2/\sqrt{40}} > 1.645$$

$$\bar{x} > 12 + 1.645\left(\frac{3.2}{\sqrt{40}}\right) = 12.8323$$

4. We will accept H_0 when $\bar{x} \leq 12.8323$

38

Calculating the Probability of a Type II Error

5. Probabilities that the sample mean will be in the acceptance region:

Values of μ	$z = \frac{12.8323 - \mu}{3.2/\sqrt{40}}$	β	$1-\beta$
14.0	-2.31	.0104	.9896
13.6	-1.52	.0643	.9357
13.2	-0.73	.2327	.7673
12.83	0.00	.5000	.5000
12.8	0.06	.5239	.4761
12.4	0.85	.8023	.1977
12.0001	1.645	.9500	.0500

39

Calculating the Probability of a Type II Error

Observations about the preceding table:

- When the true population mean μ is close to the null hypothesis value of 12, there is a high probability that we will make a Type II error.
- When the true population mean μ is far above the null hypothesis value of 12, there is a low probability that we will make a Type II error.

40

Power of the Test

- The probability of correctly rejecting H_0 when it is false is called the power of the test.
- For any particular value of μ, the power is $1 - \beta$.
- We can show graphically the power associated with each value of μ; such a graph is called a power curve.

41

Determining the Sample Size for a Hypothesis Test

$$n = (z + z_\beta)^2 \sigma^2 / (\mu_0 - \mu_a)^2$$

where

z_α = z value providing an area of α in the tail
z_β = z value providing an area of β in the tail
σ = population standard deviation
μ_0 = value of the population mean in H_0
μ_a = value of the population mean used for the Type II error

Note: In a two-tailed hypothesis test, use $z_{\alpha/2}$ not z_α

42

Relationship among α, β, and n

- Once two of the three values are known, the other can be computed.
- For a given level of significance α, increasing the sample size n will reduce β.
- For a given sample size n, decreasing α will increase β, whereas increasing α will decrease b.

43