Ch 9.1 Basics of Hypothesis Testing

Hypothesis μ, p, or σ

in statistics, is a claim or statement about a property of a population

A **significance test** is a formal procedure for comparing observed data with a claim (also called a hypothesis) whose truth we want to assess. The claim is a statement about a parameter, like the population proportion p or the population mean μ. We express the results of a significance test in terms of a probability that measures how well the data and the claim agree.

Feb 14-6:44 PM

Null Hypothesis: H_0

- Statement about value of population parameter
- Must contain condition of equality
- = $H_0: p = \#$, $\mu = \#$, $\sigma = \#$
- Test the Null Hypothesis directly
- Reject H_0 or fail to reject H_0
  False True

Feb 14-6:44 PM

Alternative Hypothesis: H_a

- Must be true if H_0 is false
- $\neq$, $<$, $>$

Feb 14-6:47 PM

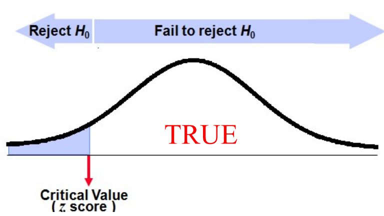
Significance Level

- denoted by α alpha
- the probability that the test statistic will fall in the critical region when the null hypothesis is actually true.
- common choices are 0.05, 0.01, and 0.10

Feb 14-6:50 PM

Critical Value

Value or values that separate the critical region (where we reject the null hypothesis) from the values of the test statistics that do not lead to a rejection of the null hypothesis



Feb 14-6:48 PM

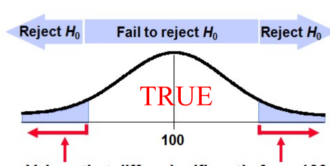
Two-tailed Test

H_0: $\mu = 100$

H_a: $\mu \neq 100$

α is divided equally between the two tails of the critical region

Means less than or greater than



Values that differ significantly from 100

Feb 14-6:54 PM

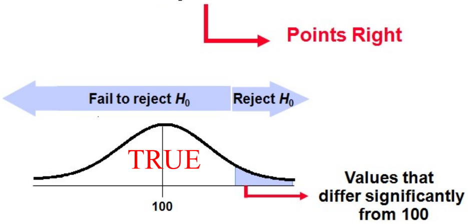
Always shade the direction the of the arrow for H_a

Right-tailed Test

H_0: μ = 100

H_a: μ > 100

H_a: population > ? parameter

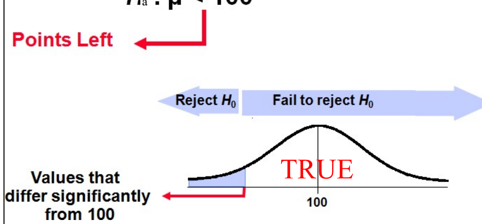


Feb 14-6:55 PM

Left-tailed Test

H_0: μ = 100

H_a: μ < 100



Feb 14-6:55 PM

Determine the null and alternative hypothesis for the following statements:

Example 1: At the Hawaii Pineapple Company, managers are interested in the size of the pineapples grown in the company's fields. Last year, the mean weight of the pineapples harvested from one large field was 31 ounces. A different irrigation system was installed in this field after the growing season. Managers wonder how this change will affect the mean weight of pineapples grown in the field this year. State appropriate hypotheses for performing a significance test. Be sure to define the parameter of interest.

H_0: $\mu = 31$
H_a: $\mu \neq 31$

Ex1c) The mean age of a college student is less than 25 years old.
H_0:
H_1:

Ex1d) The mean age of a antique car in the museum is greater than 35 years old.
H_0:
H_1:

Feb 17-1:48 PM

Ex1b) Mike is an avid golfer who would like to improve his play. A friend suggests getting new clubs and lets Mike try out his 7-iron. Based on years of experience, Mike has established that the mean distance that balls travel when his with his old 7-iron is 175 yards and a standard deviation of 15 yards. He is hoping that this new club will make his shots more consistent(less variable).

H_o: σ = 15
H_a: σ < 15

Ex

1. According to the Web site sleepdeprivation.com, 85% of teens are getting less than eight hours of sleep a night. Jannie wonders whether this result holds in her large high school. She asks an SRS of 100 students at the school how much sleep they get on a typical night. In all, 75 of the responders said less than 8 hours.

Correct Answer
(a) p = proportion of all students at Jannie's high school who get less than 8 hours of sleep at night. (b) H_0:p = 0.85 and H_a:p ≠ 0.85.

2. As part of its 2010 census marketing campaign, the U.S. Census Bureau advertised "10 questions, 10 minutes—that's all it takes." On the census form itself, we read, "The U.S. Census Bureau estimates that, for the average household, this form will take about 10 minutes to complete, including the time for reviewing the instructions and answers." We suspect that the actual time it takes to complete the form may be longer than advertised.

Correct Answer
(a) μ = true mean amount of time that it takes to complete the census form. (b) H_0:μ = 10 and H_a:μ > 10.

Jan 18-9:23 AM

P-Value Method

p-value = Shaded Region

.0244(2) = .0488

Compare the p-value to the α. If p-value ≤ α, reject H_0.

DEFINITION: *P*-value

The probability, computed assuming H_0 is true, that the statistic (such as $\hat{p}$ or $\bar{x}$) would take a value as extreme as or more extreme than the one actually observed, in the direction specified by H_a, is called the ***P*-value** of the test.

It's the probability of the sample statistic occuring in the alternative hypothesis when the null hypothesis is true.

Jan 18-9:33 AM

Example 2 Calcium is a vital nutrient for healthy bones and teeth. The National Institutes of Health(NIH) recommends a calcium intake of 1300 mg per day for teenagers. The NIH is concerned that teenagers aren't getting enough calcium. Is this true?

Researchers want to perform a test of H_0: $\mu = 1300$, H_a: $\mu < 1300$

where μ is the true mean daily calcium intake in the population of teenagers. They ask a random sample of 20 teens to record their food and drink consumption for 1 day. The researchers then compute the calcium intake for each student. Data analysis reveals that $\bar{x} = 1198$ mg and $s_x = 411$ mg. After checking that conditions were met, researchers performed a significance test and obtained a *P*-value of 0.1404.

a) Explain what it would mean for the null hypothesis to be true in this setting.

In this setting, H_0: $\mu = 1300$ says that the mean daily calcium intake in the population of teenagers is 1300 mg. If H_0 is true, then teenagers are getting enough calcium, on average.

b) Interpret the *P*-value in context.

Assuming that the mean daily calcium intake in the teen population is 1300 mg, there is a 0.1404 probability of getting a sample mean of 1198 mg or less just by chance in a random sample of 20 teens.

OR

Assuming that the mean daily calcium intake in the teen population is 1300 mg is true, then the likelihood of getting a sample of size 20 with a mean of 1198 mg or smaller is roughly 14%.

Jan 18-9:39 AM

Ex2b) Mike is an avid golfer who would like to improve his play. A friend suggests getting new clubs and lets Mike try out his 7-iron. Based on years of experience, Mike has established that the mean distance that balls travel when his with his old 7-iron is 175 yards and a standard deviation of 15 yards. He is hoping that this new club will make his shots more consistent. He hit 50 shots and a standard deviation of 13.9 yards.

$H_o: \sigma = 15$
$Ha: \sigma < 15$

If Miked performed a signficance test and obtained a p-value of .25 explain what it would mean for the null hypothesis to be true in this setting?

It means that the true standard deviation is 15 yards.

Interpret the p-value in this context.

Assuming that the true standard deviation is 15 yards, there is a .25 probability that the sample standard deviation would be 13.9 yards or less by chance alone in a sample of 50 shots with the 9-iron.

Jan 18-10:12 AM

Conclusions in Hypothesis Testing / Significance Testing

- always test the null hypothesis

1. **Reject** the H_0 when p-value < α
2. **Fail to reject** the H_0 when p-value is NOT < α

- need to formulate correct wording of final conclusion

Conclusion:

AP EXAM TIP: The conclusion to a significance test should always include three components: (1) an explicit comparison of the *P*-value to a stated significance level, (2) a decision about the null hypothesis: reject or fail to reject H_0, and (3) a statement in the context of the problem about whether or not there is convincing evidence for H_a.

Feb 14-6:55 PM

Example 3: A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery. However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. The company selects an SRS of 15 new batteries and uses them continuously until they are completely drained. The sample mean lifetime is = 33.9 hours. A significance test is performed using the hypotheses

$H_0: \mu = 30$ hours
$H_a: \mu > 30$ hours

where μ is the true mean lifetime of the new deluxe AAA batteries. The resulting *P*-value is 0.0729. What conclusion would you make for each of the following significance levels? Justify your answer.

a) $\alpha = 0.05$?

Fail to reject H_0. Because the *P*-value, 0.0729, is greater than $\alpha = 0.05$, there is not enough evidence to suggest that the mean lifetime of the new deluxe AAA batteries is greater than 30 hours, on average.

b) $\alpha = 0.10$?

Reject H_0. Because the *P*-value, 0.0729, is less than $\alpha = 0.10$, there is enough evidence to suggest that the mean lifetime of the new deluxe AAA batteries is greater than 30 hours, on average.

Jan 18-10:31 AM

Ex3b)

A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery.[2] However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. The company selects an SRS of 15 new batteries and uses them continuously until they are completely drained. The sample mean lifetime is $\bar{x} = 33.9$ hours. A significance test is performed using the hypotheses

$$H_0: \mu = 30 \text{ hours}$$
$$H_a: \mu > 30 \text{ hours}$$

where μ is the true mean lifetime of the new deluxe AAA batteries. The resulting *P*-value is 0.0729.

PROBLEM: What conclusion would you make for each of the following significance levels? Justify your answer.

(a) $\alpha = 0.10$

(b) $\alpha = 0.05$

SOLUTION:

(a) Because the *P*-value, 0.0729, is less than $\alpha = 0.10$, we reject H_0. We have convincing evidence that the company's deluxe AAA batteries last longer than 30 hours, on average.

(b) Because the *P*-value, 0.0729, is greater than $\alpha = 0.05$, we fail to reject H_0. We do not have convincing evidence that the company's deluxe AAA batteries last longer than 30 hours, on average.

Jan 18-11:01 AM

Ch 9.1 Day 2

Type I Error

- The mistake of rejecting the null hypothesis when it is true.
- α (alpha) is used to represent the probability of a type I error
- **Example:** Rejecting a claim that the mean body temperature is 98.6 degrees when the mean really does equal 98.6

Say it's false, when it really was true.

Feb 14-7:02 PM

Type 2 you say its True

Type II Error

- the mistake of failing to reject the null hypothesis when it is false.
- ß (beta) is used to represent the probability of a type II error
- **Example:** Failing to reject the claim that the mean body temperature is 98.6 degrees when the mean is really different from 98.6

Say it's true, when it really was false.

Feb 14-7:03 PM

Ex5) Give the type I and II errors for the following examples. It is helpful to identify the H_0 and H_1 first.

a) The average age for a U.S. president is 54.8 years.

H_0: $\mu = 54.8$ H_1: $\mu \neq 54.8$

Type I error:

rejecting the H_0: $\mu=54.8$, when $\mu = 54.8$

Type II error:

fail to reject H_o: $\mu = 54.8$ when $\mu \neq 54.8$

b) The mean score of an NBA basketball game is less than 100 points per game.

H_o: $\mu \geq 100$ H_1: $\mu < 100$

Type I error:

rejecting H_0: $\mu \geq 100$ when $\mu \geq 100$

Type II error:

fail to rejecting H_0: $\mu \geq 100$ when $\mu < 100$

Feb 17-3:05 PM

A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have "blemishes," the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips. To make the decision, a supervisor will inspect a random sample of potatoes from the shipment. The producer will then perform a significance test using the hypotheses

$$H_0: p = 0.08$$
$$H_a: p > 0.08$$

where p is the actual proportion of potatoes with blemishes in a given truckload.

PROBLEM: Describe a Type I and a Type II error in this setting, and explain the consequences of each.

SOLUTION: A Type I error occurs if we reject H_0 when H_0 is true. That would happen if the producer finds convincing evidence that the proportion of potatoes with blemishes is greater than 0.08 when the actual proportion is 0.08 (or less). Consequence: The potato-chip producer sends the truckload of acceptable potatoes away, which may result in lost revenue for the supplier. Furthermore, the producer will have to wait for another shipment of potatoes before producing the next batch of potato chips.

A Type II error occurs if we fail to reject H_0 when H_a is true. That would happen if the producer does not find convincing evidence that more than 8% of the potatoes in the shipment have blemishes when that is actually the case. Consequence: The producer uses the truckload of potatoes to make potato chips. More chips will be made with blemished potatoes, which may upset customers.

Jan 19-11:06 AM

Example 5 **Better Batteries**

Statistical significance

A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery.[2] However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. The company selects an SRS of 15 new batteries and uses them continuously until they are completely drained. The sample mean lifetime is $\bar{x}$ = 33.9 hours. A significance test is performed using the hypotheses

$$H_0: \mu = 30 \text{ hours}$$
$$H_a: \mu > 30 \text{ hours}$$

where μ is the true mean lifetime of the new deluxe AAA batteries. The resulting P-value is 0.0729.

1. Describe a Type I error in this setting. [redacted]

Show Answer

2. Describe a Type II error in this setting. [redacted]

Show Answer

3. Which type of error is more serious in this case? Justify your answer.

Show Answer

[redacted]

Jan 19-11:08 AM