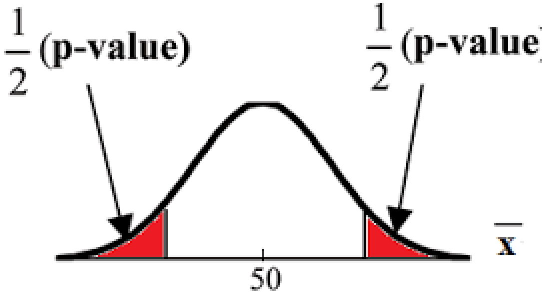
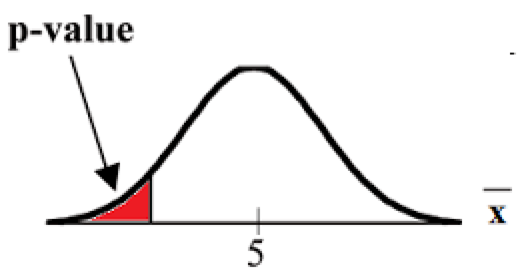
STATISTICAL INFERENCE

HYPOTHESIS TESTING

STATISTICAL INFERENCE

HYPOTHESIS TESTING:
POPULATION MEAN
POPULATION PROPORTION





WHAT WE WILL DISCUSS IN HYPOTHESIS TESTING?

- Introduction
- The Language of Hypothesis Testing
- Hypothesis Test for Population Mean
- Hypothesis Test for Population Proportion
- Type II error and the Power of Test

RECAP: INFERENTIAL STATISTICS

ESTIMATION

- "guessing" the value of the parameter
- provide the measure of the quality (reliability) of the guess

HYPOTHESIS TESTING

- making a "yes-no" decision regarding the parameter
- understand the chances of making incorrect decision

INTRODUCTION

WHY YOU NEED TO KNOW ABOUT HYPOTHESIS TESTING?

- Hypothesis testing is performed regularly in many industries.
  - Example: Companies in the pharmaceutical industry must perform many hypothesis tests on new drug products before they are deemed to be safe and effective by the federal Food and Drug Administration (FDA).

- Describe a relationship among samples
  - Does the test samples are statistically difference or not

- Hypothesis – a statement regarding a characteristics of or more populations

- Hypothesis Testing – testing the statement (hypothesis statement) using sample data regarding a characteristics of one or more populations.

- Steps in hypothesis testing:
  I. Make a statement regarding the nature of the population
  II. Collect evidence (sample data) to test the statement
  III. Analyze the data to assess the plausibility of the statement.

- cannot state 100% certainty that the statement is TRUE because sample data is used to test hypothesis.
- can only determine whether the sample data support the statement or not

THE LANGUAGE OF HYPOTHESIS TESTING

- Null Hypothesis
  - Denoted as H_o
  - A statement to be tested- The null hypothesis will be rejected only if the sample data provide substantial contradictory evidence.
    - A statement of no change, no effect or no difference and is assumed true until evidence indicates otherwise.
  - Always contain a statement of equality

- Alternative Hypothesis
  - Denoted as H_1 or H_A
  - A statement that we are trying to find evidence to support - The alternative hypothesis will be selected only if there is strong enough sample evidence to support it.
  - Always contain a statement of non equality (either not equal or less than or greater than)

TYPES OF HYPOTHESIS

Type of Hypothesis Test	Null Hypothesis	Alternative Hypothesis
Comparison of samples	The samples are the same; the samples come from the same population; the characteristics of the samples differ only because of sampling randomness	The samples are fundamentally different
Correlation	There is no significant correlation between the variables	The correlation is statistically significant
Normality	The data are normally distributed	The data are not normally distributed

- Three ways to setup the Null and the Alternative Hypothesis:

1. Equal hypothesis versus not equal hypothesis (**two tailed test**)

   H_o: parameter = some value

   H_1: parameter ≠ some value

one tailed test

2. Equal versus less than (left-tailed test)

   H_o: parameter = some value

   H_1: parameter < some value

3. Equal versus grater than (right-tailed test)

   H_o: parameter = some value

   H_1: parameter > some value

GIVING CONCLUSION IN HYPOTHESIS TESTING

When the null hypothesis is rejected, we say that there is sufficient evidence to support the statement in the alternative hypothesis.

When the null hypothesis is not rejected (accepted),we say that there is not sufficient evidence to support the statement in the alternative hypothesis.

Example:
When making comparisons between 2 sample means, there are 2 possibilities

Null hypothesis is true

Accept Null Hypothesis

Reject Alternative Hypothesis

Null hypothesis is false

Reject Null hypothesis

Accept Alternative hypothesis

FORMING HYPOTHESIS

- How to form hypothesis?
  a) Identify the parameter to be tested

  b) Determine the status quo value of the parameter – gives the null hypothesis

  c) Determine the statement that reflect what we are trying to gather evidence – gives the alternative hypothesis

look for key phrase

FORMULATING THE HYPOTHESIS

Example:

In today's economy, university students often work many hours to help pay for the high costs of a college education. Suppose a university in the Midwest is considering changing its class schedule to accommodate students working long hours. The registrar has stated a change is needed because the mean number of hours worked by undergraduate students at the university is more than 20 hours per week.

Steps to solution:

1. **Determine the population parameter of interest.**

Population parameter is mean, μ number of hours worked by undergraduate students

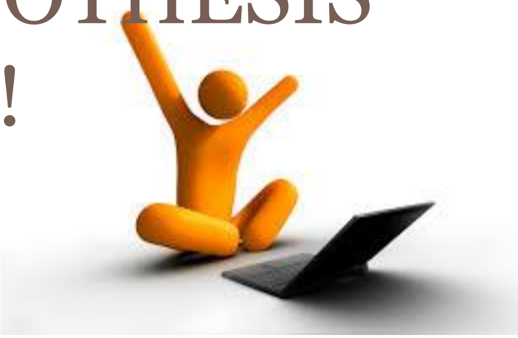
2. **Identify the hypothesis of interest**

The registrar has made a claim that the mean hours worked "is more than 20 hours " per week. The hypothesis will not be declared true unless the sample data strongly indicate that it is true. Thus, the burden of proof is placed on the registrar to justify her claim that the mean is greater than 20 hours.

3. **Formulate the null and the alternative hypothesis**

$$H_0: \mu = 20 \text{ hours per week}$$
$$H_1: \mu > 20 \text{ hours per week (claim)}$$

FORMULATING THE HYPOTHESIS LETS PRACTICE!!



Example 1:

The Frito-Lay Company produces several snack and food products that are sold throughout the United States and around the world. The company uses an automatic filling machine to fill the sacks with the desired weight. For instance, when the company is running potato chips on the fill line, the machine is set to fill the sacks with 20 ounces. Thus, if the machine is working properly, the mean fill will be 20 ounces. Each hour, a sample of sacks is collected and weighed, and the technicians determine whether the machine is still operating correctly or whether it needs adjustment.

Example 2:

The director of a state agency believes that the average starting salary for clerical employees in the state is less than $30,000 per year. To test her hypothesis, she has collected a simple random sample of 100 starting clerical salaries from across the state and found that the sample mean is $29,750. State the appropriate null and alternative hypotheses.

FORMING HYPOTHESIS

- Form the hypothesis
  a) The GSK pharmaceutical company has just developed a new antibiotics for children. Two percent of children taking competing antibiotics experience headaches as a side effect. A researcher for the Food and Drug Administration wishes to know if the percentage of children taking the new antibiotic who experience headaches as a side effect is more than 2%.

  b) The Blue Book price of used 3 year old Chevy Corvette is $86,012. Grant wonders if the mean price of a used 3 year old Chevy Corvette in the Miami metropolitan area is different form $86,012.

  c) The standard deviation of the contents in a 64 ounce bottle of detergent using in old filling machine is 0.23 ounce. The manufacturer wants to know if a new filling machine has less variability.

STATISTICAL ERRORS

- understand the chances of making incorrect decision

Example:

If a drug company’s hypothesis tests for a new drug incorrectly conclude that the drug is safe when in fact it is not. The company’s customers may become ill or even die as a result.

On the other hand, if the hypothesis tests incorrectly determined that the drug was not safe, thus, most likely the potentially useful and safe drug would most likely not be made available to people who need it.

FOUR OUTCOME FROM HYPOTHESIS TESTING

1. Reject the null hypothesis when the alternative hypothesis is true. This decision would be **correct**.

2. Do not reject the null hypothesis when the null hypothesis is true. This decision would be **correct**.

3. Reject the null hypothesis when the null hypothesis is true. This decision would be **incorrect**. This type of error is called a **Type I error**.

4. Do not reject the null hypothesis when the alternative hypothesis is true. This decision would be **incorrect**. This type of error is called a **Type II error**.

STATISTICAL ERRORS

Figure: Two type of statistical error that can be made in Hypothesis Testing

		Reality	
		H_0 Is True	H_1 Is True
Conclusion	Do Not Reject H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

Type I error: α or False Positive (FP)
Type II error: β or False Negative (FN)

Confusion matrix in data mining

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

STATISTICAL ERRORS: TYPE I ERROR

- The probability of a Type I error is controlled by the decision maker by the choice of significance level, α.
  - called as 'the size of the test'

- If type I error is fixed at 5 % ($\alpha = 0.05$) , it means that there are about 5 chances in 100 that we will reject H0 when the reality the H0 is true.

STATISTICAL ERRORS: TYPE II ERROR

- Type II error is committed when we fail to believe a truth.
- A type II error occurs when one rejects the alternative hypothesis (fails to reject the null hypothesis) when the alternative hypothesis is true.
- The rate of the type II error is denoted by the Greek letter β (beta) and related to the power of a test (which equals 1-β).

SIGNIFICANCE LEVEL AND CRITICAL VALUE

- We need to select the *cutoff* point that is as the separation between rejecting and not rejecting the null hypothesis.

- Significance level (or level of significance):
  - The maximum allowable probability of committing a Type I statistical error.
  - The probability is denoted by the symbol α.

SIGNIFICANCE LEVEL AND CRITICAL VALUE

- Critical value:
  - Having chosen a significance level, α, the decision maker then must calculate the corresponding cutoff point.
    - The value corresponding to a significance level that determines those test statistics that lead to rejecting the null hypothesis and those that lead to a decision not to reject the null hypothesis.
  - Then it is compared to the calculated test statistic to determine whether to reject or accept the null hypothesis.

- If the **absolute value** of your test statistic is greater than the **critical value**, you can declare statistical significance and reject the null hypothesis.

How to find critical value:
You need to refer z-table (standard normal table):
For example:

- The level of significance, $\alpha = 0.05$
- If it is two tail (lower and upper) test, thus, $0.05/2 = 0.025$
  - From table locate the value of 0.025 and the critical value = 1.96

- If it is one-tail (lower or upper) test,
  - From table locate the value of 0.05 and the critical value = 1.645

ONE TAILED OR TWO TAILED

Two type of hypothesis tests:

- **one-tailed test**: the entire rejection region is located in one tail (either upper tail or lower tail) of sampling distribution.

Example:

$H_0 : \mu = 25$ days

$H_A : \mu > 25$ days

the entire region is located in the upper tail & the null hypothesis will be rejected only when the sample mean falls in the extreme upper tail of the sample distribution

- **two-tailed test**: the entire rejection region is split into two tails (upper tail and lower tail) of sampling distribution.

Example:

$H_0 : \mu = 25$ days

$H_A : \mu \neq 25$ days

The null hypothesis will be rejected only when the sample mean falls in the extreme upper tail or extremely small (lower tail) of the sample distribution

TEST STATISTICS AND DECISION RULES

- Test statistics:
  - A function of the sampled observations that provides a basis for testing a statistical hypothesis.
    - The calculated *z*-value is an example of test statistics.

- Decision rules: You can use three equivalent approaches: ** does not make no difference as each method yields the same conclusion
  1. calculate a *test statistics value* and compare it to the critical value from the distribution (z-distribution, student's t distribution and etc)
  2. calculate a sample mean, $\bar{x}$ (or proportion mean, $\widehat{p}$) and compare it to the critical value, $\bar{x}_\alpha$ (or $\hat{p}_\alpha$)
  3. use the method of p-value approach

AN EXAMPLE OF CALCULATED TEST STATISTICS

z-Test Statistic for Hypothesis Tests for μ,σ Known

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where:

$\bar{x}$ = Sample mean
μ = Hypothesized value for the population mean
σ = Population standard deviation
n = Sample size

$\bar{x}_\alpha$ for Hypothesis Tests for μ,σ Known

$$\bar{x}_\alpha = \mu + z_\alpha \frac{\sigma}{\sqrt{n}}$$

where:

μ = Hypothesized value for the population mean
z_α = Critical value from the standard normal distribution
σ = Population standard deviation
n = Sample size

STATISTICAL HYPOTHESIS TESTING

Single Population:

- Single pop. mean, the sample mean ($\bar{x}$) is used to test the hypothesis: either σ known or σ unknown
- Single pop. proportion, the sample proportion ($\hat{p}$) is used to test the hypothesis

Two Population:

- Two pop. mean (with independent sample), the sample mean ($\bar{x}$ for sample 1 and sample 2) is used to test the hypothesis: either (σ_1 and σ_2) known or (σ_1 and σ_2) unknown
- Two pop. mean (with paired sample), the sample mean ($\bar{x}$ $\bar{x}$ for sample 1 and sample 2) is used to test the hypothesis: either (σ_1 and σ_2) known or (σ_1 and σ_2) unknown
- Two pop. proportion, the sample statistics ($\hat{p}$) is used to test the hypothesis

HYPOTHESIS TESTING FOR SINGLE POPULATION MEAN

HYPOTHESIS TESTING FOR SINGLE POPULATION MEAN

- 2 Scenarios can be considered as:
  - either standard deviation, σ known or
  - standard deviation, σ unknown
- Steps in hypothesis testing:
  1. Determine the null and alternative hypotheses.
  2. Select a level of significance, α.
  3. Compute the test statistic.
  4. Compare the critical value to the test statistic.
  5. State the conclusion.

SCENARIO 1: standard deviation, σ known

z-Test Statistic for Hypothesis Tests for μ,σ Known

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where:

$\bar{x}$ = Sample mean
μ = Hypothesized value for the population mean
σ = Population standard deviation
n = Sample size

$\bar{x}_\alpha$ for Hypothesis Tests for μ,σ Known

$$\bar{x}_\alpha = \mu + z_\alpha \frac{\sigma}{\sqrt{n}}$$

where:

μ = Hypothesized value for the population mean
z_α = Critical value from the standard normal distribution
σ = Population standard deviation
n = Sample size

DECISION RULES AND TEST STATISTICS

one-tailed test with lower tail

- **Approach 1**: calculate a z-value and compare it to the critical value, z_α
  - if the -z-value ≥ critical value, $-z_\alpha$ therefore accept H_o
  - if the -z-value < critical value, $-z_\alpha$ therefore reject H_o

- **Approach 2**: calculate a sample mean, $\bar{x}$ and compare it to the critical value, $\bar{x}_\alpha$
  - if the -$\bar{x}$ ≥ critical value, -$\bar{x}_\alpha$ therefore accept H_o
  - if the -$\bar{x}$ < critical value, -$\bar{x}_\alpha$ therefore reject H_o

- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated –(p-value) ≥ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated – (p-value) < probability in the rejection region, α therefore reject H_o

DECISION RULES AND TEST STATISTICS

one-tailed test with upper tail

- **Approach 1**: calculate a *z*-value and compare it to the critical value, z_α
  - if the *z*-value >critical value, z_α therefore reject H_o
  - if the *z*-value $\leq$ critical value, z_α therefore accept H_o

- **Approach 2**: calculate a sample mean, $\bar{x}$ and compare it to the critical value, $\bar{x}_\alpha$
  - if the $\bar{x}$ > critical value, $\bar{x}_\alpha$ therefore reject H_o
  - if the $\bar{x} \leq$ critical value,$\bar{x}_\alpha$ therefore accept H_o

- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated p-value $\geq$ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated p-value < probability in the rejection region, α therefore reject H_o

DECISION RULES AND TEST STATISTICS

TWO-tailed test

- **Approach 1**: calculate a z-value and compare it to the critical value, z_α
  - if the $-z$ value < critical value, $-z_{\alpha/2}$ and critical value, $z_{\alpha/2} > z$ value therefore reject H_o
  - if the $-z$ value $\geq$ critical value, $-z_{\alpha/2}$ and critical value, $z_{\alpha/2} \leq z$ value therefore accept H_o
- **Approach 2**: calculate a sample mean, $\bar{x}$ and compare it to the critical value, $\bar{x}_\alpha$
  - if the $-\bar{x}$ < critical value, $-\bar{x}_{\alpha/2}$ and critical value, $\bar{x}_{\alpha/2} > \bar{x}$ value therefore reject H_o
  - if the $-\bar{x} \geq$ critical value, $-\bar{x}_{\alpha/2}$ and critical value, $\bar{x}_{\alpha/2} \leq \bar{x}$ value therefore accept H_o
- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated p-value ≥ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated p-value < probability in the rejection region, α therefore reject H_o

HYPOTHESIS TESTING FOR SINGLE POPULATION

Example 1

The ABX Specialist Hospital in United State, performs many knee replacement surgery procedures each year. Recently, research physicians at that hospital have developed a surgery process they believe will reduce the average patient recovery time. The hospital board will not recommend the new procedure unless there is substantial evidence to suggest that it is better than the existing procedure. Records indicate that the current mean recovery rate for the standard procedure is 142 days, with a standard deviation of 15 days. To test whether the new procedure actually results in a lower mean recovery time, the procedure was performed on a random sample of 36 patients and the $\bar{x}$ is assume 140.2 days ~ 140 days. The researchers wish to test the hypothesis using a 0.05 level of significance.

Steps to solution:

Hypothesis claim: new procedure of knee replacement gives lower mean recovery time ($\mu < 142$ days)

Formulate the null and the alternative hypothesis

$$H_0 : \mu = 142 \text{ days}$$

$$H_A : \mu < 142 \text{ days (claim)}$$

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM

σ = *15 days*

Construct the rejection region and decision rule
This is one-tailed test with lower tail (left hand) of sampling distribution. With 0.05 level of significance, the critical value, $-z_{0.05} = -1.645$

Compute the test statistic
This problem will use z-test statistics.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{140.2 - 142}{\frac{15}{\sqrt{36}}} = -0.72$$

The decision rule: If $-z < -1.645$, then reject H_o, otherwise accept H_o
Since $-0.72 > -1.645$, therefore H_o is accepted.

Draw a conclusion
There is no sufficient evidence to conclude that the new knee replacement procedure results in a shorter average recovery period. Thus, ABX Specialist Hospital will not be able to recommend the new procedure on the grounds that it reduces recovery time.

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM (LETS PRACTICE)

Example 1 (one-tailed hypothesis test for μ, σ known)

The Testing Center in Southern California creates standardized exams for a variety of quantitative disciplines, including business statistics. Recently the Testing Center received complaints from faculty who have used its latest business statistics test saying the mean time required to complete the exam exceeds the advertised mean of 40 minutes. Before responding, employees at the Testing Center plan to test this claim using an alpha level equal to 0.05 and a random sample size of 100 business statistics students. Suppose that the sample of 100 students produced a sample mean of 43.5 minutes. Based on previous studies, suppose that the population standard deviation is known to be $\sigma = 8$ minutes.

Find the Answer:

a) **Hypothesis claim:**
b) **Formulate the null and the alternative hypothesis**
c) **Construct the rejection region and decision rule**
d) **Compute the test statistic**
e) **Identify the decision rule**
f) **Draw a conclusion**

Example 1 (one-tailed hypothesis test for μ, σ known) - ANSWER

The Testing Center in Southern California creates standardized exams for a variety of quantitative disciplines, including business statistics. Recently the Testing Center received complaints from faculty who have used its latest business statistics test saying the mean time required to complete the exam exceeds the advertised mean of 40 minutes. Before responding, employees at the Testing Center plan to test this claim using an alpha level equal to 0.05 and a random sample size of n 100 business statistics students. Suppose that the sample of 100 students produced a sample mean of 43.5 minutes. Based on previous studies, suppose that the population standard deviation is known to be $\sigma = 8$ minutes.

The Answer:

Hypothesis claim: The new claim states the mean time required to complete the exam exceeds (>) the advertised mean of 40 minutes

Formulate the null and the alternative hypothesis

$$H_0 : \mu = 40 \text{ minutes}$$

$$H_A : \mu > 40 \text{ minutes (claim)}$$

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM

σ = *8 minutes*

Construct the rejection region and decision rule

This is one-tailed test with upper tail (right hand) of sampling distribution. With 0.05 level of significance, the critical value, $z_{0.05}$= 1.645

Compute the test statistic

This problem will use $\bar{x}$ statistics.

$$\bar{x}_\alpha = \mu + z_\alpha \frac{\sigma}{\sqrt{n}} = 40 + 1.645 \frac{8}{\sqrt{100}} = 41.32$$

The decision rule: if the $\bar{x}$ < critical value, $\bar{x}_\alpha$ therefore accept H_o, otherwise reject

Since 43.5 > 41.32, therefore H_o is rejected.

Draw a conclusion

There is sufficient evidence to conclude that the mean time required to complete the exam exceeds the advertised time of 40 minutes. The Testing Center will likely want to modify the exam to shorten the average completion time.

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM (LETS PRACTICE)

Example 2

The ABA Corporation is a wood products company with lumber, plywood, and paper plants in several areas of Malaysia. At its ABX subsidiaries, plywood plant, the company makes plywood used in residential and commercial building. One product made at the plant is 3/8-inch plywood which must have a mean thickness of 0.375 inches. The standard deviation, σ, is known to be 0.05 inch. Before sending a shipment to customers, the managers test whether they are meeting the 0.375 inch requirements by selecting a random sample of n 100 sheets of plywood and collecting thickness measurements. Suppose the sample mean for the random sample of 100 measurement is 0.378 inch. Construct the statistical hypothesis test with significance level, $\alpha = 0.05$ and help the manager to state the decision.

Find the Answer:

a) **Hypothesis claim:**
b) **Formulate the null and the alternative hypothesis**
c) **Construct the rejection region and decision rule**
d) **Compute the test statistic**
e) **Identify the decision rule**
f) **Draw a conclusion**

Example 2 (two-tailed hypothesis test for μ, σ known) - ANSWER

The ABA Corporation is a wood products company with lumber, plywood, and paper plants in several areas of Malaysia. At its ABX subsidiaries, plywood plant, the company makes plywood used in residential and commercial building. One product made at the plant is 3/8-inch plywood which must have a mean thickness of 0.375 inches. The standard deviation, σ, is known to be 0.05 inch. Before sending a shipment to customers, the managers test whether they are meeting the 0.375 inch requirements by selecting a random sample of n 100 sheets of plywood and collecting thickness measurements. Suppose the sample mean for the random sample of 100 measurement is 0.378 inch. Construct the statistical hypothesis test with significance level, $\alpha = 0.05$ and help the manager to state the decision.

The Answer:

Hypothesis claim: The 3/8-inch plywood which must have a mean thickness of 0.375 inches.

Formulate the null and the alternative hypothesis

$$H_o : \mu = 0.375 \text{ inches (claim)}$$

$$H_A : \mu \neq 0.375 \text{ inches}$$

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM

σ = *0.05 inches*

Construct the rejection region and decision rule

This is two-tailed test. With 0.05 level of significance, the critical value, for upper tail $z_{\alpha/2} = z_{0.05/2} = z_{0.025} = 1.96$ and for lower tail $-z_{\alpha/2}$ = - $z_{0.05/2} = -z_{0.025}$ = -1.96

Compute the test statistic

This problem will use z-test statistics.

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.378 - 0.375}{\frac{0.05}{\sqrt{100}}} = 0.6$$

The decision rule: if the $-z$ value < critical value, $-z_{\alpha/2}$ and critical value, $z_{\alpha/2}$ > z value therefore reject H_o , otherwise accept H_o

Since - 1.96 < - 0.60 and 0.60 < 1.96 (two-tailed test), therefore H_o is accepted.

Draw a conclusion

There is sufficient evidence that the product met the 0.375 inch requirements. Therefore, they can continue with the shipment of the plywood.

SCENARIO 2: standard deviation, σ unknown

Testing hypotheses about a mean for **standard deviation, σ unknown** follows the same logic as testing a hypothesis for standard deviation, σ known .
The only difference is that **it use Student's t-distribution**, rather than the normal distribution.

t-Test Statistic for Hypothesis Tests for μ,σ unknown

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where:

$\bar{x}$ = Sample mean
μ = Hypothesized value for the population mean
s = Sample standard deviation, $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
n = Sample size

with degree of freedom: *n-1*

STATISTICAL HYPOTHESIS TESTING: EXAMPLE OF PROBLEM

Example 1 (hypothesis test for μ, σ unknown)

The Dairy Fresh Ice Cream uses a filling machine for its 64-ounce cartons. There is some variation in the actual amount of ice cream that goes into the carton. The machine can go out of adjustment and put a mean amount either less or more than 64 ounces in the cartons. To monitor the filling process, the production manager selects a simple random sample of 16 filled ice cream cartons each day. The test is conducted using α = 0.05From the hypothesis test, the manager wants to know either the machine is in adjustment or not. Since, the standard deviation is unknown, the manager have run the pilot sample and the data is as follows:

62.7	64.7	64.0	64.5	64.6	65.0	64.4	64.2
64.6	65.5	63.6	64.7	64.0	64.2	63.0	63.6

Steps to solution:

Hypothesis claim: The machine can go out of adjustment and put a mean amount either less or more than 64 ounces in the cartons.

Formulate the null and the alternative hypothesis

$H_0 : \mu = 64$ ounce (machine is in adjustment)

$H_A : \mu \neq 64$ ounce (machine is out of adjustment)

Construct the rejection region and decision rule

This is two-tailed test. With 0.05 level of significance. The degree of freedom for t-distribution is: n-1= 16-1=15 degree of freedom. Therefore, $t = \pm 2.131$

Decision rule: if the $-t_{calc}$ value < critical value, -*t* and critical value, t > t_{calc} value therefore reject H_o , otherwise accept H_o

Compute the test statistic

This problem will use t-test statistics.

The sample mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{1,027.3}{16} = 64.2$$

The sample standard deviation is

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = 0.72$$

The *t*-test statistic, using Equation 9.3, is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{64.2 - 64}{\frac{0.72}{\sqrt{16}}} = 1.11$$

The decision rule: because -t= -1.11 is greater than -2.131 and 1.11 is not greater than 2.131, therefore H_o is accepted.

Draw a conclusion

Based on these sample data, the company does not have sufficient evidence to conclude that the filling machine is out of adjustment.

HYPOTHESIS TESTING FOR POPULATION PROPORTION

HYPOTHESIS TEST FOR SINGLE POPULATION PROPORTION

- Previously: hypothesis testing about single population mean.
- Why need hypothesis test for proportion?
  - Example: A production manager might consider the proportion of defective items produced on an assembly line to determine whether the line should be restructured.
  - Example: A life insurance salesperson's performance assessment might include the proportion of existing clients who renew their policies.
- The basic concepts of hypothesis testing for proportions are the same as for means.

HYPOTHESIS TEST FOR SINGLE POPULATION PROPORTION

$\bar{p}_\alpha$- critical value for Proportions (approach 2)

$$\bar{p}_\alpha = p + z_\alpha \sqrt{\frac{p(1-p)}{n}}$$

where:
$\bar{p}$ = sample proportion
n = sample size
p = hypothesized population proportion
Z_α= z critical value

z test-statistics for Proportions (approach 1)

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Approach 2 : calculate $\bar{p}$ value and compare it to the critical value, $\bar{p}_\alpha$

- if the $\bar{p}$ value ≥ critical value,$\bar{p}_\alpha$therefore reject H_o
- if the $\bar{p}$ value < critical value,$\bar{p}_\alpha$therefore accept (do not reject) H_o

Approach 3: p-value (same as testing for population mean)

STATISTICAL HYPOTHESIS TESTING: DECISION RULES AND TEST STATISTICS

one-tailed test with lower tail

- **Approach 1**: calculate a *z*-value and compare it to the critical value, z_α
  - if the -*z*-value ≥ critical value, $-z_\alpha$ therefore accept H_o
  - if the -*z*-value < critical value, $-z_\alpha$ therefore reject H_o

- **Approach 2** : calculate $\bar{p}$ value and compare it to the critical value, $\bar{p}_\alpha$
  - if the -$\bar{p}$ value ≥ critical value,$\bar{p}_\alpha$therefore accept H_o
  - if the -$\bar{p}$ value < critical value,$\bar{p}_\alpha$therefore reject H_o

- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated –(p-value) ≥ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated – (p-value) < probability in the rejection region, α therefore reject H_o

STATISTICAL HYPOTHESIS TESTING: DECISION RULES AND TEST STATISTICS

one-tailed test with upper tail

- **Approach 1**: calculate a z-value and compare it to the critical value, z_α
  - if the z-value $\geq$ critical value, z_α therefore reject H_o
  - if the z-value $<$ critical value, z_α therefore accept H_o
- **Approach 2** : calculate $\bar{p}$ value and compare it to the critical value, $\bar{p}_\alpha$
  - if the $\bar{p}$ value $\geq$ critical value,$\bar{p}_\alpha$therefore reject H_o
  - if the $\bar{p}$ value $<$ critical value,$\bar{p}_\alpha$therefore accept H_o

- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated p-value $\geq$ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated p-value $<$ probability in the rejection region, α therefore reject H_o

DECISION RULES AND TEST STATISTICS

TWO-tailed test

- **Approach 1**: calculate a z-value and compare it to the critical value, z_α
  - if the $-z$ value < critical value, $-z_{\alpha/2}$ and critical value, $z_{\alpha/2} > z$ value therefore reject H_o
  - if the $-z$ value ≥ critical value, $-z_{\alpha/2}$ and critical value, $z_{\alpha/2} \le z$ value therefore accept H_o
- **Approach 2**: calculate a sample mean, $\bar{x}$ and compare it to the critical value, $\bar{x}_\alpha$
  - if the $-\bar{p}$ < critical value, $-\bar{p}_{\alpha/2}$ and critical value, $\bar{p}_{\alpha/2} > \bar{p}$ value therefore reject H_o
  - if the $-\bar{p} \ge$ critical value, $-\bar{p}_{\alpha/2}$ and critical value, $\bar{p}_{\alpha/2} \le \bar{p}$ value therefore accept H_o
- **Approach 3**: use the method of p-value approach and compare it to the probability in the rejection region, α
  - if the calculated p-value ≥ probability in the rejection region, α therefore accept (do not reject) H_o
  - if the calculated p-value < probability in the rejection region, α therefore reject H_o

EXAMPLE OF PROBLEM (LETS PRACTICE)

Example

The NBA started a new professional basketball league called the Developmental League (D League), where players who were not on NBA rosters could fine-tune their skills in hopes of getting called up to the NBA. The teams in this league are privately owned but connected to NBA teams. One of the D-League's teams is considering increasing the season ticket prices for basketball games. The marketing manager is **concerned** that some people will terminate their ticket orders if this change occurs. If more than 10% of the season ticket orders would be terminated, the marketing manager does not want to implement the price increase. To test this, a random sample of ticket holders is surveyed and asked what they would do if the prices were increased. The random sample of n = 100 season ticket holders showed that 14 would cancel their ticket orders if the price change were implemented. The hypothesis testing is carried with significance level, $\alpha = 0.05$

The Solution:

Hypothesis claim: population of proportion that season ticket holder would terminate their ticket orders if price increasing

Formulate the null and the alternative hypothesis

$$H_0 : p = 0.10$$
$$H_A : p > 0.10 \text{ (claim)}$$

Construct the rejection region and decision rule

This is one-tailed test. With 0.05 level of significance, the critical value, for upper tail $z_\alpha = z_{0.05} = 1.645$

Compute the test statistic

This problem will use z-test statistics for population proportion.

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

need to calculate $\bar{p} = \frac{x}{n}$, where the random sample of n = 100 season ticket holders showed that 14 would cancel their ticket orders if the price change were implemented.

Therefore, $\bar{p} = \frac{14}{100} = 0.14$

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.14 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{100}}} = \frac{0.04}{0.03} = 1.33$$

The decision rule: if the z_{cal} < critical value, z_{α} therefore accept H_o, otherwise reject

Since, 1.33< 1.645, therefore accept H_o

Draw a conclusion

Based on the sample data, the marketing manager does not have sufficient evidence to conclude that more than 10% of the season ticket holders will cancel their ticket orders

What happen if the decision rule is based from $\bar{p}_{\alpha}$- critical value (p-value) for proportions. Will it gives the same conclusion?

EXAMPLE OF PROBLEM (LETS PRACTICE)

Example

The two major college entrance exams that a majority of colleges accept for admission are the SAT and ACT. ACT looked at historical records and established 22 as the minimum ACT math score for a student to be considered prepared for college mathematics. An official with the Illinois State Department of Education wonders whether less than half of the students in her state are prepared for College Algebra. She obtains a simple random sample of 500 records of students who have taken the ACT and finds that 219 are prepared for college mathematics (that is, scored at least 22 on the ACT math test). Does this represent significant evidence that less than half of Illinois students who have taken the ACT are prepared for college mathematics upon graduation from a high school? Use the $\alpha = 0.05$ level of significance.

The Solution:

Hypothesis claim: We want to determine if the sample evidence shows that less than half of the students are prepared for college mathematics. Symbolically, we represent this as $p < 0.5$.

Formulate the null and the alternative hypothesis

$$H_0 : p = 0.5$$
$$H_A : p < 0.5 \text{(claim)}$$

Construct the rejection region and decision rule

This is one-tailed test. With 0.05 level of significance, the critical value, for lower tail $-z_\alpha = -z_{0.05} = -1.645$

Compute the test statistic

This problem will use z-test statistics for population proportion.

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

need to calculate $\bar{p} = \frac{x}{n}$, where the random sample of n = 500 season ticket holders showed that 219 is prepared for college mathematics.

Therefore, $\bar{p} = \frac{219}{500} = 0.438$

$$z = \frac{\bar{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.438-0.5}{\sqrt{\frac{0.5(1-0.5)}{500}}} = -2.77$$

The decision rule: if the $-z_{cal}$ > critical value, $-z_{\alpha}$ therefore accept H_o , otherwise reject

Since, $-2.77 < -1.645$, therefore reject H_o

Draw a conclusion

There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that fewer than half of the Illinois students are prepared for college mathematics. In other words, the data suggest less than a majority of the students in the state of Illinois who take the ACT are prepared for college mathematics.

HYPOTHESIS TESTING FOR TWO POPULATION MEANS:

Independent Samples

Paired Samples

STATISTICAL INFERENCE

HYPOTHESIS TESTING FOR TWO POPULATION MEANS:

Independent Samples

HYPOTHESIS TESTS FOR TWO POPULATION MEANS USING INDEPENDENT SAMPLES

- Why need to know? - require to test whether two populations have equal means, or whether one population mean is larger (smaller) than another.
- The hypothesis testing between the means will cover:
  1. The population standard deviation are known and the samples are independent
  2. The population standard deviation are unknown and the samples are independent

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE KNOWN, USING INDEPENDENT SAMPLES

z- test statistic for $\mu_1 - \mu_2$ when σ_1 and σ_2 are known, using independent samples

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where

$(\mu_1 - \mu_2)$ = Hypothesized difference in population means

Decision rule:

- ➢if the z-value ≥ critical value, z_α therefore accept (do not reject) H_o
- ➢if the z-value < critical value, z_α therefore reject H_o

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE **KNOWN**, USING INDEPENDENT SAMPLES

Problem:

AMCO Drilling Equipment, Inc., is a company that makes drilling equipment for the oil and gas industry. One item that the company makes is a coupling for use on natural gas drills. AMCO has two machines that make these couplings. It is well established that the standard deviation for the coupling's diameter made by machine 1 is 0.025 inches and the standard deviation for machine 2 is 0.034 inches. These are known values. However, the company is interested in determining whether there is a difference in the average diameters made by these machine. The company wishes to know whether machine 2 also provides couplings with higher average diameters than the average diameter produced by machine 1. If the test determines that machine 2 has a larger average diameter than machine 1, the managers will have maintenance attempt to adjust the diameters downward or they will replace the machine. The test will be conducted using $\alpha = 0.05$.

The manager have select simple random samples of 100 couplings from the two populations (machine 1 and machine 2 productions) and compute the sample means. The means computed from the samples are $\bar{x}_1 = 0.501$ inches and $\bar{x}_2 = 0.509$ inches.

Steps to solution:

1. Specify the population parameter of interest.

This μ_1-μ_2 the difference in the two population means.
The samples are independent because the diameters of couplings made by one machine can in no way influence the diameter of the couplings made by the other machine.

2. Formulate the appropriate null and alternative hypotheses.

We are interested in determining whether the mean diameter for machine 2 exceeds that for machine 1.

$$H_0 : \mu_1 = \mu_2$$
$$H_A : : \mu_1 < \mu_2 \text{ (claim)}$$

3. Determine critical value based from significance level and state the decision rule.

This is one-tailed test. With 0.05 level of significance, the critical value, for upper tail $-z_\alpha = -z_{0.05} = -1.645$

The decision rule: if the z < critical value, z_α therefore reject H_0, otherwise accept

Alternatively, you can state the decision rule in terms of a *p*-value,
if the *p-value* < α therefore reject H_0, otherwise accept H_0

4. Compute the test statistic

This problem will use zstatistics.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(0.501 - 0.509) - 0}{\sqrt{\frac{0.025^2}{100} + \frac{0.034^2}{100}}} = -1.90$$

The decision rule: Since -1.90 < - 1.645, therefore H_o is rejected.

5. Identify the summary for the problem studied

There is statistical evidence to conclude that the couplings made by machine 2 have a larger mean diameter than those made by machine 1. Thus, AMCO managers need to take action to modify the mean diameters from machine 2 or replace it.

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE **UNKNOWN**, USING INDEPENDENT SAMPLES

- When testing the two population mean when the population standard deviation is unknown and the sample sizes are small, the critical value is a *t*-value from the *t*-distribution.

- The following assumptions hold:
  - Each populations has a normal distributed.
  - The two population variance are equal $(\sigma_1{}^2, \sigma_2{}^2)$.
  - The samples are independent.

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE **UNKNOWN**, USING INDEPENDENT SAMPLES

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad df = n_1 + n_2 - 2$$

where:

$\bar{x}_1$ and $\bar{x}_2$ = Sample means from populations 1 and 2

$\mu_1 - \mu_2$ = Hypothesized difference between population means

n_1 and n_2 = Sample sizes from the two populations

s_p = Pooled standard deviation (see Equation 10.4)

t-Test Statistic for $\mu_1 - \mu_2$ When σ_1 and σ_2 Are Unknown and Assumed Equal, Independent Samples

$$s_p = \sqrt{\frac{(n_1 - 1)s^2{}_1 + (n_2 - 1)s^2{}_2}{n_1 + n_2 - 2}}$$

ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN σ_1 AND σ_2 ARE UNKNOWN, USING INDEPENDENT SAMPLES

Problem:

A recent Associated Press news story out of Brussels, Belgium, indicated the European Union was considering a probe of computer makers after consumers complained that they were being overcharged for ink cartridges. Companies such as Canon, Hewlett-Packard, and Epson are the printer market leaders and make most of their printer related profits by selling replacement ink cartridges. Suppose an independent test agency wishes to conduct a test to determine whether name-brand ink cartridges generate more color pages on average than competing generic ink cartridges. The test is conducted with $\alpha = 0.05$. A simple random sample of 10 users was selected, and the users were given a name-brand cartridge. A second sample of 8 users was given generic cartridges. Both groups used their printers until the ink ran out. The number of pages printed was recorded. The samples are independent because the pages printed by users in one group did not in any way influence the pages printed by users in the second group. The means computed from the samples are

Name-brand cartridge	Generic cartridge
$\bar{x}_1 = 322.5$ pages	$\bar{x}_1 = 298.3$pages
$S_1 = 48.3$ pages	$S_2 = 53.3$ pages

Steps to solution:

1. **Define the population parameter of interest**

We are interested in determining whether the mean number of pages printed by name-brand cartridges (population 1) exceeds the mean pages printed by generic cartridges (population 2).

2. **Formulate the hypothesis null and alternative**

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 > \mu_2 \text{ (claim)}$$

3. **Construct the rejection region**

This is one-tailed test. With 0.05 level of significance, the critical value of *t*-value from the t distribution, $t = (n_1+n_2-2)$ *degree of freedom*, thus $10 + 8 - 2 = 16$ degree of freedom. From the t-table, the critical value t is for upper tail $t_\alpha = t_{0.05} = 1.7459$

The decision rule: if the t > critical value, t_α therefore reject H_o, otherwise accept H_o

Alternatively, you can state the decision rule in terms of a *p*-value, if the *p-value* > α therefore reject H_o, otherwise accept H_o

Steps to solution:

4. Determine the test statistics

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

The pooled standard deviation is

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(10-1)48.3^2 + (8-1)53.3^2}{10+8-2}} = 50.55$$

Then the t-test statistic is

$$t = \frac{(322.5 - 298.3) - 0.0}{50.55\sqrt{\frac{1}{10} + \frac{1}{8}}} = 1.0093$$

5. Reach the decision

t = 1.0093 < $t_{0.05}$= 1.7459 . Therefore, acept H_o

6. Identify the summary for the problem studied

Based on these sample data, there is insufficient evidence to conclude that the mean number of pages produced by name-brand ink cartridges exceeds the mean for generic cartridges.

STATISTICAL INFERENCE

HYPOTHESIS TESTING FOR TWO POPULATION MEANS:

Paired Samples

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE UNKNOWN, USING PAIRED SAMPLES

- Recap:
  - What is paired samples? : Samples that are selected in such a way that values in one sample are matched with the values in the second sample for the purpose of controlling for extraneous factors.
  - Another term for paired samples is dependent samples.
  - Why paired samples? – to control for any variation in a sample. Example: different cars (and drivers), different painter and etc

TESTING FOR $\mu_1 - \mu_2$ WHEN σ_1 AND σ_2 ARE **UNKNOWN**, USING PAIRED SAMPLES

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}, \qquad df = (n-1)$$

where:

$$\bar{d} = \text{Mean paired difference} = \frac{\sum d}{n}$$

$$\mu_d = \text{Hypothesized population mean paired difference}$$

$$s_d = \text{Sample standard deviation for paired differences} = \sqrt{\frac{\sum(d - \bar{d})^2}{n-1}}$$

$$n = \text{Number of paired values in the sample}$$

***t*-Test Statistic for Paired Sample Test**

ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN σ_1 AND σ_2 ARE UNKNOWN, USING INDEPENDENT SAMPLES

Problem:

Referring to previous example, suppose the experiment regarding ink cartridges is conducted differently. Instead of having different samples of people use name-brand and generic cartridges, the test is done using paired samples. This means that the same people will use both types of cartridges, and the pages printed in each case will be recorded. The test is conducted with $\alpha = 0.01$. Six randomly selected people have agreed to participate and will be tested on the name-brand and generic cartridges. The following data and paired differences were observed:

Printer User	Name-brand cartridge	Generic cartridge
1	306	300
2	256	260
3	402	357
4	299	286
5	306	290
6	257	260

Steps to solution:

1. **Define the population parameter of interest**

We are interested in determining whether name-brand cartridges produce more printed pages, on average, than generic cartridges, so we would expect the paired difference to be positive. We assume that the paired differences are normally distributed..

2. **Formulate the hypothesis null and alternative**

$$H_o : \mu_d = 0.0$$
$$H_A : \mu_d > 0.0 \text{ (claim)}$$

3. **Construct the rejection region**

This is one-tailed test. With 0.01 level of significance, the critical value of *t*-value from the t distribution, *t = (n_1-1) degree of freedom*, thus $6 - 1 = 5$ degree of freedom. From the t-table, the critical value t is for upper tail $t_\alpha = t_{0.01} = 3.3649$

The decision rule: if the t > critical value, t_α therefore reject H_o, otherwise accept H_o

Alternatively, you can state the decision rule in terms of a *p*-value, if the *p-value* > α therefore reject H_o, otherwise accept H_o

Steps to solution:

4. Determine the test statistics

The mean paired difference is:

$$\bar{d} = \frac{\sum d}{n} = 12.17$$

The standard deviation for the paired difference is:

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n - 1}} = 18.02$$

The test statistics:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{12.17 - 0.00}{\frac{18.02}{\sqrt{6}}} = 1.6543$$

Printer User	Name-brand cartridge	Generic cartridge	d
1	306	300	6
2	256	260	-4
3	402	357	45
4	299	286	13
5	306	290	16
6	257	260	-3

5. Reach the decision

$t = 1.6543 < t_{0.01} = 3.3649$. Therefore, accept H_o

6. Identify the summary for the problem studied

Based on these sample data, there is insufficient evidence to conclude that name-brand ink cartridges produce more pages on average than generic brands.

EXERCISE

The director of a state agency believes that the average starting salary for clerical employees in the state is less than \$30,000 per year. To test her hypothesis, she has collected a simple random sample of 100 starting clerical salaries from across the state and found that the sample mean is \$29,750.

a) State the appropriate null and alternative hypotheses.
b) Assuming the population standard deviation is known to be \$2,500 and the significance level for the test is to be 0.05, what is the critical value (stated in dollars)?
c) Referring to your answer in part b, what conclusion should be reached with respect to the null hypothesis?

Assume that the sports page of your local newspaper reported that 65% of males over the age of 17 in the United States would skip an important event such as a birthday party or an anniversary dinner to watch their favorite professional sports team play. A random sample of 676 adult males over the age of 17 in the Dallas-Fort Worth market reveals that 507 would be willing to skip an important event to watch their favorite team play. Given the results of the survey, can you conclude that the proportion of adult males who would skip an important event to watch their favorite team play is greater in the Dallas Fort Worth area than in the nation as a whole? Conduct your test at the $\alpha = 0.01$ level of significance.

The marketing manager for a major retail grocery chain is wondering about the location of the stores' dairy products. She believes that the mean amount spent by customers on dairy products per visit is higher in stores where the dairy section is in the central part of the store compared with stores that have the dairy section at the rear of the store. To consider relocating the dairy products, the manager feels that the increase in the mean amount spent by customers must be at least 25 cents. To determine whether relocation is justified, her staff selected a random sample of 25 customers at stores where the dairy section is central in the store. A second sample of 25 customers was selected in stores with the dairy section at the rear of the store. The following sample results were observed:

Central Dairy	Rear Dairy
$\bar{x}_1 = \$3.74$	$\bar{x}_2 = \$3.26$
$s_1 = \$0.87$	$s_2 = \$0.79$

Conduct a hypothesis test with a significance level of 0.05 to determine if the manager should relocate the dairy products in those stores displaying their dairy products in the rear of the store.

Production engineers at Megamind.co believe that a modified layout on its assembly lines might increase average worker productivity (measured in the number of units produced per hour). However, before the engineers are ready to install the revised layout officially across the entire firm's production lines, they would like to study the modified line's effects on output. The following data represent the average hourly production output of 12 randomly sampled employees before and after the line was modified:

Employee	1	2	3	4	5	6	7	8	9	10	11	12
Before	49	45	43	44	48	42	46	46	49	42	46	44
After	49	46	48	50	46	50	45	46	47	51	51	49

At the 0.05 level of significance, can the production engineers conclude that the modified (after) layout has increased average worker productivity?

Suppose as part of a national study of economic competitiveness a marketing research firm randomly sampled 200 adults between the ages of 27 and 35 living in metropolitan Seattle and 180 adults between the ages of 27 and 35 living in metropolitan Minneapolis. Each adult selected in the sample was asked, among other things, whether they had a college degree. From the Seattle sample 66 adults answered yes and from the Minneapolis sample 63 adults answered yes when asked if they had a college degree. Based on the sample data, can we conclude that there is a difference between the population proportions of adults between the ages of 27 and 35 in the two cities with college degrees? Use a level of significance of 0.01 to conduct the hypothesis test.

THE END OF ITEM 2