Two-sample Tests of Hypothesis

Chapter 11

Dr. Sultana Parveen, Professor, IPE Dept, BUET

Learning Objectives

LO1 Test a hypothesis that two independent population means with known population standard deviations are equal.

LO2 Carry out a hypothesis test that two population proportions are equal.

LO3 Conduct a hypothesis test that two independent population means are equal assuming equal but unknown population standard deviations.

LO4 Conduct a test of a hypothesis that two independent population means are equal assuming unequal but unknown population standard deviations.

LO5 Explain the difference between dependent and independent samples.

LO6 Carry out a test of hypothesis about the mean difference between paired and dependent observations.

Comparing two populations – Some Examples

1. Is there a difference in the mean value of residential real estate sold by male agents and female agents in south Florida?
2. Is there a difference in the mean number of defects produced on the day and the afternoon shifts at Kimble Products?
3. Is there a difference in the mean number of days absent between young workers (under 21 years of age) and older workers (more than 60 years of age) in the fast-food industry?
4. Is there is a difference in the proportion of Ohio State University graduates and University of Cincinnati graduates who pass the state Certified Public Accountant Examination on their first attempt?
5. Is there an increase in the production rate if music is piped into the production area?

LO1 Test a hypothesis that two independent population means with known population standard deviations are equal.

Comparing Two Population Means

- No assumptions about the shape of the populations are required.
- The samples are from independent populations.
- The formula for computing the test statistic (*z*) is:

Use if sample sizes > 30 or if σ_1 and σ_2 are known

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Use if sample sizes > 30 and if σ_1 and σ_2 are unknown

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

EXAMPLE

The U-Scan facility was recently installed at the Byrne Road Food-Town location. The store manager would like to know if the **mean checkout time** using the standard checkout method **is longer** than using the U-Scan. She gathered the following sample information. The time is measured from when the customer enters the line until their bags are in the cart. Hence the time includes both waiting in line and checking out.

Customer Type	Sample Mean	Population Standard Deviation	Sample Size
Standard	5.50 minutes	0.40 minutes	50
U-Scan	5.30 minutes	0.30 minutes	100

Step 1: State the null and alternate hypotheses.

(keyword: "longer than")

H_0: $\mu_S \leq \mu_U$

H_1: $\mu_S > \mu_U$

Step 2: Select the level of significance.

The .01 significance level is stated in the problem.

LO1

Example 1 *continued*

Step 3: Determine the appropriate test statistic.

Because both population standard deviations are known, we can use *z-distribution* as the test statistic

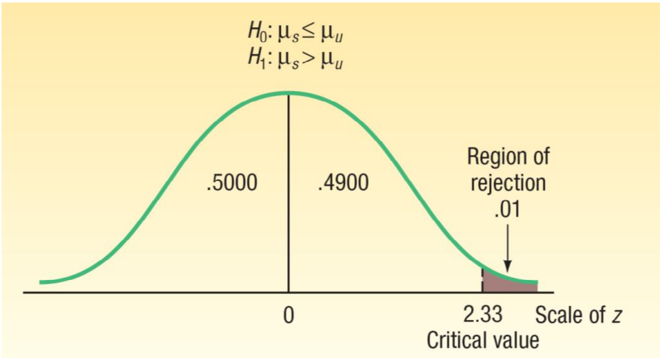
Step 4: Formulate a decision rule.

Reject H_0 if $Z > Z_\alpha$

$Z > 2.33$

Step 5: Compute the value of *z* and make a decision

$$z = \frac{\bar{X}_s - \bar{X}_u}{\sqrt{\frac{\sigma_s^2}{n_s} + \frac{\sigma_u^2}{n_u}}} = \frac{5.5 - 5.3}{\sqrt{\frac{0.40^2}{50} + \frac{0.30^2}{100}}} = \frac{0.2}{0.064} = 3.13$$



The computed value of 3.13 is larger than the critical value of 2.33. Our decision is to reject the null hypothesis. The difference of .20 minutes between the mean checkout time using the standard method is too large to have occurred by chance. We conclude the U-Scan method is faster.

LO3 Conduct a hypothesis test that two independent population means are equal assuming equal but unknown population standard deviations.

Comparing Population Means with Unknown Population Standard Deviations (the Pooled *t*-test)

The *t* distribution is used as the test statistic if one or more of the samples have less than 30 observations. The required assumptions are:

1. Both populations must follow the normal distribution.
2. The populations must have equal standard deviations.
3. The samples are from independent populations.

Finding the value of the test statistic requires two steps.

1. Pool the sample standard deviations.
2. Use the pooled standard deviation in the formula.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

EXAMPLE

Owens Lawn Care, Inc., manufactures and assembles lawnmowers that are shipped to dealers throughout the United States and Canada. Two different procedures have been proposed for mounting the engine on the frame of the lawnmower. The question is: **Is there a difference in the mean time to mount the engines on the frames of the lawnmowers?**

To evaluate the two methods, it was decided to conduct a time and motion study. A sample of five employees was timed using the Welles method and six using the Atkins method. The results, in minutes, are shown below:

Welles (minutes)	Atkins (minutes)
2	3
4	7
9	5
3	8
2	4
	3

Is there a difference in the mean mounting times? Use the .10 significance level.

Comparing Population Means with Unknown Population Standard Deviations (the Pooled *t*-test) - Example

Step 1: State the null and alternate hypotheses.

(Keyword: "Is there a *difference*")

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Step 2: State the level of significance.

The 0.10 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

Because the population standard deviations are not known but are assumed to be equal, we use the pooled *t*-test.

Step 4: State the decision rule.

Reject H_0 if $t > t_{\alpha/2, n_1+n_2-2}$ or $t < -t_{\alpha/2, n_1+n_2-2}$

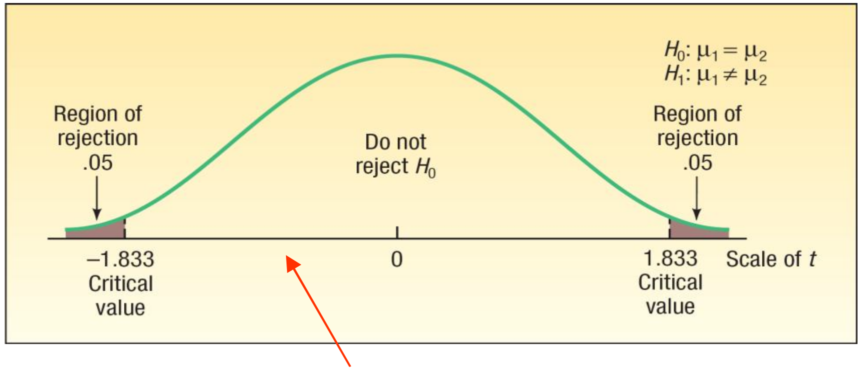
$t > t_{.05,9}$ or $t < -t_{.05,9}$

$t > 1.833$ or $t < -1.833$

Step 5: Compute the value of t and make a decision

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(5 - 1)(2.9155)^2 + (6 - 1)(2.0976)^2}{5 + 6 - 2} = 6.2222$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{4.00 - 5.00}{\sqrt{6.2222\left(\frac{1}{5} + \frac{1}{6}\right)}} = -0.662$$



The decision is **not to reject the null hypothesis**, because **-0.662 falls in the region between -1.833 and 1.833**.

We conclude that there is no difference in the mean times to mount the engine on the frame using the two methods

LO5 Carry out a test of hypothesis about the mean difference between paired and dependent observations.

Two-Sample Tests of Hypothesis: Dependent Samples

Dependent samples are samples that are paired or related in some fashion.

For example:

- If you wished to buy a car you would look at the *same* car at two (or more) *different* dealerships and compare the prices.
- If you wished to measure the effectiveness of a new diet you would weigh the dieters at the start and at the finish of the program.

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Where
$\bar{d}$ is the mean of the differences
s_d is the standard deviation of the differences
n is the number of pairs (differences)

EXAMPLE
Nickel Savings and Loan wishes to compare the two companies it uses to appraise the value of residential homes. Nickel Savings selected a sample of 10 residential properties and scheduled both firms for an appraisal. The results, reported in $000, are shown on the table (right).

At the .05 significance level, can we conclude there is a difference in the mean appraised values of the homes?

Home	Schadek	Bowyer
1	235	228
2	210	205
3	231	219
4	242	240
5	205	198
6	230	223
7	231	227
8	210	215
9	225	222
10	249	245

Hypothesis Testing Involving Paired Observations - Example

Step 1: State the null and alternate hypotheses.

H_0: $\mu_d = 0$

H_1: $\mu_d \neq 0$

Step 2: State the level of significance.

The .05 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

We will use the *t*-test

Step 4: State the decision rule.

Reject H_0 if

$t > t_{\alpha/2,\ n-1}$ or $t < -t_{\alpha/2,n-1}$

$t > t_{.025,9}$ or $t < -t_{.025,\ 9}$

$t > 2.262$ or $t < -2.262$

LO5

Hypothesis Testing Involving Paired Observations - Example

Step 5: Compute the value of *t* and make a decision

The computed value of t (3.305) is greater than the higher critical value (2.262), so our decision is to reject the null hypothesis.

We conclude that there is a difference in the mean appraised values of the homes.

Home	Schadek	Bowyer	Difference, d	$(d - \bar{d})$	$(d - \bar{d})^2$
1	235	228	7	2.4	5.76
2	210	205	5	0.4	0.16
3	231	219	12	7.4	54.76
4	242	240	2	−2.6	6.76
5	205	198	7	2.4	5.76
6	230	223	7	2.4	5.76
7	231	227	4	−0.6	0.36
8	210	215	−5	−9.6	92.16
9	225	222	3	−1.6	2.56
10	249	245	4	−0.6	0.36
			46	0	174.40

$$\bar{d} = \frac{\Sigma d}{n} = \frac{46}{10} = 4.60$$

$$s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}} = \sqrt{\frac{174.4}{10 - 1}} = 4.402$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{4.6}{4.402/\sqrt{10}} = 3.305$$