Chapter 10: Two Population Hypothesis Tests

Department of Mathematics
Izmir University of Economics

Week 13-14
2014-2015

In this chapter, we will focus on hypothesis testing procedures for

- **difference between two population means**.

Our discussion follows the development in Chapter 9, based on comparing two populations resulted in Chapter 8.

Throughout this chapter the **test statistics** is:

$$\text{test statistics} = \frac{(\bar{x} - \bar{y}) - D_0}{\text{standard error}},$$

where **standard error** changes depending on whether the populations are

- either dependent or
- independent with known variances or
- independent with unknown and assumed equal variances or
- independent with unknown but not assumed equal variances.

Dependent samples

Here, we

- test the difference between means of two **related** populations,
- use differences between paired values:

$$d_i = x_i - y_i,$$

- and continue with the usual hypothesis testing procedures using these d_i's.

- Upper-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \leq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y > D_0$$

with test statistic

$$t = \frac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t > t_{n-1,\alpha}.$$

- Lower-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \geq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y < D_0$$

with test statistic

$$t = \frac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{n-1,\alpha}.$$

- Two-sided hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0$$

$$\text{against } H_1 : \mu_x - \mu_y \neq D_0$$

with test statistic

$$t = \frac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{n-1,\frac{\alpha}{2}} \text{ or } t > t_{n-1,\frac{\alpha}{2}}.$$

Example: Rental car prices per gallon were sampled at eight major airports. Data for Hertz and National car rental companies is given as:

Airport	Hertz	National	Airport	Hertz	National
BL	1.55	1.56	JFK	1.72	1.51
CO'H	1.62	1.59	LaG	1.67	1.50
LA	1.72	1.78	OC	1.68	1.77
Mi	1.65	1.49	Du	1.52	1.41

Use $\alpha = 0.05$ to test the hypothesis of no difference between the population mean prices per gallon for the two companies.

Independent samples with known σ^2s

Here, we

- test the difference between means of two independent populations
- with known population variances: σ_x^2 and σ_y^2.

- Upper-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \leq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y > D_0$$

with test statistic

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } z > z_\alpha.$$

- Lower-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \geq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y < D_0$$

with test statistic

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } z < -z_\alpha.$$

- Two-sided hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0$$

$$\text{against} \quad H_1 : \mu_x - \mu_y \neq D_0$$

with test statistic

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } z < -z_{\frac{\alpha}{2}} \text{ or } z > z_{\frac{\alpha}{2}}.$$

Example: One reason for wage differentials between men and women is the differential of their work experience. Given data of experiences,

Men	Women
$n_x = 100$	$n_y = 85$
$\bar{x} = 14.9$	$\bar{y} = 10.5$
$\sigma_x = 5.2$	$\sigma_y = 3.8$

test the hypothesis that men tend to have at least 4.5 more years of experience than women.

Independent samples with unknown σ^2s - Assuming equal

This one is a more realistic model because the samples are random and independent.

Here, we

- test the difference between means of two independent populations
- with unknown population variances σ_x^2 and σ_y^2
- but we assume that they are **equal**: $\sigma_x^2 = \sigma_y^2$.

So, we define (as previous), a *pooled sample variance:*

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \text{ with } v = n_x + n_y - 2.$$

- Upper-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \leq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y > D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t > t_{v,\alpha}.$$

- Lower-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \geq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y < D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{v,\alpha}.$$

- Two-sided hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0$$

$$\text{against } H_1 : \mu_x - \mu_y \neq D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{v,\frac{\alpha}{2}} \text{ or } t > t_{v,\frac{\alpha}{2}}.$$

Example: The following results are for independent random samples taken from two populations:

Sample 1	Sample 2
$n_x = 20$	$n_y = 30$
$\bar{x} = 22.5$	$\bar{y} = 20.1$
$s_x = 2.5$	$s_y = 4.8$

It is claimed that the population means are the same. Test this hypothesis assuming that population variances are equal.

Independent samples with unknown σ^2s - Not assuming equal

Again, we have independent and random samples.

Here, we

- test the difference between means of two independent populations
- with unknown population variances σ_x^2 and σ_y^2
- and we do **not** assume they are **equal**: $\sigma_x^2 \neq \sigma_y^2$.

Recall that, we need to choose the *degrees of freedom* appropriately:

$$v = \frac{\left[\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right]^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}}$$

and we always round v down!

- Upper-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \leq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y > D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t > t_{v,\alpha}.$$

- Lower-tail hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0 \text{ or } H_0 : \mu_x - \mu_y \geq D_0$$

$$\text{against } H_1 : \mu_x - \mu_y < D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{v,\alpha}.$$

- Two-sided hypothesis test:

$$\text{Test } H_0 : \mu_x - \mu_y = D_0$$

$$\text{against} \quad H_1 : \mu_x - \mu_y \neq D_0$$

with test statistic

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

and decision rule

$$\text{Reject } H_0 \text{ if } t < -t_{v,\frac{\alpha}{2}} \text{ or } t > t_{v,\frac{\alpha}{2}}.$$

Example: According to customer satisfaction survey results, the average ratings about two consultants of having different experience levels are claimed to be equal.

Consultant A (10 years of experience)	Consultant B (1 year of experience)
$n_x = 16$	$n_y = 10$
$\bar{x} = 6.82$	$\bar{y} = 6.25$
$s_x = 0.64$	$s_y = 0.75$

Assuming unequal population variances, test the given hypothesis.

Exercises

Example: Suppose you wish to compare a new method of teaching reading to "slow learners" to the current standard method. You decide to base this comparison on the results of a reading test given at the end of a learning period of 6 months. Of a random sample of 22 slow learners, 10 are taught by the new method and 12 are taught by the standard method. All 22 children are taught by qualified instructors under similar conditions for a 6-month period. The results of the reading test at the end of this period are given

Reading Test Scores for Slow Learners							
New Method				Standard Method			
80	80	79	81	79	62	70	68
76	66	71	76	73	76	86	73
70	85			72	68	75	66

Test the null hypothesis that there is no difference between methods against the alternative hypothesis that new method is better. We assume that population variances are equal with $\alpha = 0.10$.

Example: The U.S. Department of Transportation provides the number of miles that residents of the 75 largest metropolitan areas travel per day by car. Suppose that for a simple random sample of 50 Buffalo residents the mean is 22.5 miles a day and the standard deviation is 8.4 miles per day, and for an independent simple random sample of 40 Boston residents the mean is 18.6 miles per day and the standard deviation is 7.4 miles a day. Test the hypothesis that Buffalo residents do not travel more miles than Boston residents.

Example: Safegate Foods, Inc., is redesigning the checkout lanes in its supermarkets throughout the country and is considering two designs. Tests on customer checkout times conducted at two stores where the two new systems have been installed result in the following summary of data.

System A	**System B**
$n_1 = 120$	$n_2 = 100$
$\bar{x} = 4.1$ minutes	$\bar{y} = 3.4$ minutes
$\sigma_x = 2.2$ minutes	$\sigma_y = 1.5$ minutes

Test at 0.02 significance level to determine whether the population mean checkout times of the two systems differ. Which system is preferred?