Dr. Saeed A. Dobbah Alghamdi

Department of Statistics

Faculty of Sciences

Building 90, Office 26F41

King Abdulaziz University

http://saalghamdy.kau.edu.sa

Dr. Saeed's Website

Statistics for Business & Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams

Statistical Inference about Means and Proportions with Two Populations

Chapter 10

Learning Objectives

- **LO1** Construct and interpret confidence intervals and hypothesis tests for the difference between two population means, given independent samples from the two populations
- **LO2** Construct and interpret confidence intervals and hypothesis tests for the difference between two population means, given matched samples from the two populations.
- **LO3** Construct and interpret confidence intervals and hypothesis tests for the difference between two population proportions, given independent samples from the two populations.

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. Average lifetimes of two different brand of tires might be compared to see whether there is any difference in tread wear.
5. Test scores of the same students in Accounting and Statistics.
6. 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?
7. Is there an increase in the production rate if music is piped into the production area?

Hypothesis Tests Forms

- Let us consider hypothesis test forms about the difference between two population means μ_1 and μ_2. There are three forms as follows:

Ho: $\mu_1 = \mu_2$ *or* Ho: $\mu_1 - \mu_2 = 0$ *or* Ho: $\mu_1 - \mu_2 = D_o$ Ha: $\mu_1 \neq \mu_2$ *or* Ha: $\mu_1 - \mu_2 \neq 0$ *or* Ha: $\mu_1 - \mu_2 \neq D_o$
Ho: $\mu_1 \geq \mu_2$ *or* Ho: $\mu_1 - \mu_2 \geq 0$ *or* Ho: $\mu_1 - \mu_2 \geq D_o$ Ha: $\mu_1 < \mu_2$ *or* Ha: $\mu_1 - \mu_2 < 0$ or Ha: $\mu_1 - \mu_2 < D_o$
Ho: $\mu_1 \leq \mu_2$ *or* Ho: $\mu_1 - \mu_2 \leq 0$ *or* Ho: $\mu_1 - \mu_2 \leq D_o$ Ha: $\mu_1 > \mu_2$ *or* Ha: $\mu_1 - \mu_2 > 0$ or Ha: $\mu_1 - \mu_2 > D_o$

- Let us consider hypothesis test forms about the difference between two population proportions p_1 and p_2. There are three forms as follows:

Ho: $p_1 = p_2$ *or* Ho: $p_1 - p_2 = 0$ *or* Ho: $p_1 - p_2 = D_o$ Ha: $p_1 \neq p_2$ *or* Ha: $p_1 - p_2 \neq 0$ *or* Ha: $p_1 - p_2 \neq D_o$
Ho: $p_1 \geq p_2$ *or* Ho: $p_1 - p_2 \geq 0$ *or* Ho: $p_1 - p_2 \geq D_o$ Ha: $p_1 < p_2$ *or* Ha: $p_1 - p_2 < 0$ or Ha: $p_1 - p_2 < D_o$
Ho: $p_1 \leq p_2$ *or* Ho: $p_1 - p_2 \leq 0$ *or* Ho: $p_1 - p_2 \leq D_o$ Ha: $p_1 > p_2$ *or* Ha: $p_1 - p_2 > 0$ or Ha: $p_1 - p_2 > D_o$

Two Population Means–z test

- No assumptions about the shape of the populations are required.
- The samples are from independent populations with known populations standard deviations σ_1 and σ_2 or with samples sizes n of at least 30.
- The interval estimate of the difference between two population means μ_1 and μ_2 is

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- The test statistics for hypothesis tests about the difference between two population means μ_1 and μ_2 is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Not: use samples standard deviations S_1 and S_2 if populations standard deviations σ_1 and σ_2 are unknown when samples sizes 30 or more.

Two Population Means–z test Example

Example

The U-Scan facility was recently installed at the Byrne Road Food-Town location. 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

The store manager would like to know the 99% confidence interval estimate of the difference between the mean of standard and u-scan checkout methods and if the **mean checkout time** using the standard checkout method **is longer** than using the U-Scan at 1% significance level.

Two Population Means–z test Example

Step 1: State the null and alternate hypotheses.

H_0: $\mu_S \leq \mu_U$ H_1: $\mu_S > \mu_U$ **or** H_0: $\mu_S - \mu_U \leq 0$ H_1: $\mu_S - \mu_U > 0$

Step 2: Select the level of significance.

The 0.01 significance level is stated in the problem.

Step 3: Determine the appropriate test statistic.

z-test since both population standard deviations are known

Step 4: Formulate a decision rule.

Reject H_0 if P-value ≤ 0.01

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

The p-value 0.0009 is less than1% indicating the rejection of H_0. Thus, there is enough evidence to support the claim that the difference between the mean checkout time using the standard method is too large to have occurred by chance. Thus, the U-Scan method is faster.

Point Estimate of the difference: $\bar{x}_s$ - $\bar{x}_u$=5.5 - 5.3 = 0.2

99% Confidence Interval: (0.0351, 0.3649)

Two Population Means–t test-Unequal Variances

- The samples are from independent populations with unknown populations standard deviations σ_1 and σ_2 and samples sizes n less than 30.
- The two standard deviations σ_1 and σ_2 are assumed not equal, $\sigma_1 \neq \sigma_2$.
- The interval estimate of the difference between two population means μ_1 and μ_2 is

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- The test statistics for hypothesis tests about the difference between two population means μ_1 and μ_2 is

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Two Population Means–t test Example

Example

Personnel in a consumer testing laboratory are evaluating the absorbency of paper towels. They wish to compare a set of store brand towels to a similar group of name brand ones. For each brand they dip a ply of the paper into a tub of fluid, allow the paper to drain back into the vat for two minutes, and then evaluate the amount of liquid the paper has taken up from the vat.
A random sample of 9 store brand paper towels absorbed the following amounts of liquid in milliliters: 8, 8, 3, 1, 9, 7, 5, 5, 12
An independent random sample of 12 name brand towels absorbed the following amounts in milliliters: 12, 11, 10, 6, 8, 9, 9, 10, 11, 9, 8, 10

Use the 0.05 significance level to construct a confidence interval for the difference in the mean amount of liquid absorbed between the two types of paper towels and test if the difference differ from zero.

Since the values of the populations' variances are unknown, we first test for the equality of the two variances.

$H_0: \sigma_1^2 = \sigma_2^2$ $\qquad H_1: \sigma_1^2 \neq \sigma_2^2$

The p-value for the F-test for equality of variance is 0.0313 indicating unequal variances for the two groups.

F-test for equality of variance			
	11.03	variance: Store	
	2.63	variance: Name	
	4.20	F	
	.0313	p-value	

Two Population Means–t test Example

Step 1: State the null and alternate hypotheses.

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

Step 2: State the level of significance.

The 0.05 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

We will use unequal variances *t*-test

Step 4: State the decision rule.

Reject H_0 if p-value ≤ 0.05

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

The p-value is 0.0329 which is less than 0.05 indicating the rejection of H_0. Thus, there is enough evidence to support the claim that the mean absorption rate for the two towels is not the same.

Point Estimate of the difference: $\bar{x}_b$ - $\bar{x}_s$=9.417 – 6.444 = 2.973

95% Confidence Interval: (0.294, 5.650)

Two Population Means–t test-Equal Variances

- The samples are from independent populations with unknown populations standard deviations σ_1 and σ_2 and samples sizes n less than 30.
- The two standard deviations σ_1 and σ_2 are assumed equal, $\sigma_1 = \sigma_2$, pool the samples standard deviations

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

- The interval estimate of the difference between two population means μ_1 and μ_2 is

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- The test statistics for hypothesis tests about the difference between two population means μ_1 and μ_2 is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Two Population Means–t test Example

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 here:

Welles: 2, 4, 8, 3, 2 Atkins: 3, 7, 5, 8, 4, 3

Use the 0.01 significance level to construct a confidence interval for the difference in the mean mounting times between the two methods and test if the difference differ from zero.

Since the populations variances are unknown, we first test for the equality of the two variances.
The p-value for the F-test for equality of variance is 0.7045 indicating equal variances for the two groups.

F-test for equality of variance			
	6.20	variance: W	
	4.40	variance: A	
	1.41	F	
	.7045	p-value	

Two Population Means–t test Example

Step 1: State the null and alternate hypotheses.

H_0: $\mu_A = \mu_W$ H_1: $\mu_A \neq \mu_W$

Step 2: State the level of significance.

The 0.01 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

Use the pooled t-test since the population standard deviations are not known but are assumed to be equal.

Step 4: State the decision rule.

Reject H_0 if p-value ≤ 0.01

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

The decision is not to reject the null hypothesis since the p-value 0.4074 is greater than 0.01. Thus, there is not enough evidence to support the claim that there is difference in the mean times to mount the engine on the frame using the two methods.

Point Estimate of the difference: $\bar{x}_A$ - $\bar{x}_W$=5 – 3.8 = 1.2

99% Confidence Interval: (-3.287, 5.687)

Two Population Means–Matched Samples

- **Matched or Dependent samples** are samples that are paired or related in some manners.
- For example:
  - If you wished to buy a car you would look for 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.
- The interval estimate of the difference between the two dependent population means of before μ_b and after μ_a is

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

- The formula for the test statistic is

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

Where

$\bar{d}$ is the mean of the differences

μ_d is hypothesized mean difference (usually zero)

s_d is the standard deviation of the differences

n is the number of pairs (differences)

Two Dependent Population Means–Example

Example

Nickel Savings and Loan employs two firms, Schadek Appraisals and Bowyer Real Estate, to appraise the value of the real estate properties on which it makes loans. To review the consistency of the two appraisal firms, 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 following table.
At the 0.05 significance level, construct confidence interval for the differences and can we conclude there is a difference in the mean appraised values of the homes?

Step 1: State the null and alternate hypotheses.

$H_0: \mu_d = 0 \qquad H_1: \mu_d \neq 0$

Step 2: State the level of significance.

The 0.05 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 p-value ≤ 0.05

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

Two Dependent Population Means–Example

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

The p-value 0.0092 is less than 0.05 indicating the rejection of the null hypothesis. Thus, there is enough evidence to support the claim that there is a difference in the mean appraised values of the homes between the two firms.

Point Estimate of the difference: $\bar{x}_s$ - $\bar{x}_b$= 226.8 – 222.2 = 4.6

95% Confidence Interval: (1.451, 7.749)

Two-Sample Tests about Proportions

- We investigate whether two samples came from populations with an equal proportion of successes.
- The interval estimate of the difference between the two population proportions p_1 and p_2 is

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \quad ; \quad \bar{p}_i = \frac{X_i}{n_i}$$

- Under the null assumption is true as an equality, the two samples are pooled using the pooled estimate of population proportion formula.

$$p_c = \frac{X_1 + X_2}{n_1 + n_2}$$

where:

X_1 is the number possessing the trait in the first sample.
X_2 is the number possessing the trait in the second sample.
n_1 is the number of observations in the first sample.
n_2 is the number of observations in the second sample.

- The test statistics for hypothesis tests about the difference between the two population proportions p_1 and p_2 is

$$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{p_c(1-p_c)}{n_1} + \frac{p_c(1-p_c)}{n_2}}}$$

Two-Sample Tests about Proportions-Example

Example

Manelli Perfume Company recently developed a new fragrance that it plans to market under the name Heavenly. A number of market studies indicate that Heavenly has very good market potential. The Sales Department at Manelli is particularly interested in whether **there is a difference in the proportions of younger and older** women who would purchase Heavenly if it were marketed at 5% significance level. Samples are collected from each of these independent groups. Each sampled woman was asked to smell Heavenly and indicate whether she likes the fragrance well enough to purchase a bottle.

The random sample of 100 young women revealed 19 liked the Heavenly fragrance well enough to purchase it. Similarly, a sample of 200 older women revealed 62 liked the fragrance well enough to make a purchase. Let p_1 refer to the young women and p_2 to the older women.

Two-Sample Tests about Proportions-Example

Step 1: State the null and alternate hypotheses.

$H_0: P_1 = P_2$ $\qquad$ $H_1: P_1 \neq P_2$

Step 2: Select the level of significance.

The 0.05 significance level is stated in the problem.

Step 3: Determine the appropriate test statistic.

We will use the z-distribution

Step 4: Formulate the decision rule.

Reject H_0 if P-value ≤ 0.05

5: Select a sample and make a decision

The p-value is 0.0273 smaller than 0.05 indicating the rejection of H_0. Thus, there is enough evidence to support the claim that the proportion of young women who would purchase Heavenly is not equal to the proportion of older women who would purchase Heavenly.

Point Estimate of the difference: $\bar{p}_1 - \bar{p}_2$= 0.19 – 0.31 = -0.12

95% Confidence Interval: (-0.2201, -0.0199)

Summary

Independent simple random samples Samples selected from two populations in such a way that the elements making up one sample are chosen independently of the elements making up the other sample.

Matched samples Samples in which each data value of one sample is matched with a corresponding data value of the other sample.

Pooled estimator of *p* An estimator of a population proportion obtained by computing a weighted average of the point estimators obtained from two independent samples.