

In this chapter, we will learn how to test a claim comparing parameters from two populations. To conduct inference about two population parameters, we must first determine the sampling distribution of the difference of two parameters.

Recall: Two samples are **independent** if the sample values selected from one population are not related to or somehow paired with the sample values selected from the other population.

### 9-2 Inferences about two proportions:

In this section, we learn how to use a z-test to test the difference between two population proportions. Suppose a simple random sample of size  $n_1$  is taken from a population where  $x_1$  of the individuals have a specified characteristic, and a sample of size  $n_2$  is independently taken from a different population where  $x_2$  of the individuals taken from a different population where  $x_2$  of the individuals have a specified characteristic.

The **sampling distribution of the difference between the two proportions**  $\hat{p}_1 - \hat{p}_2$ , is approximately normal, with mean  $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ . When testing a hypothesis made about two population proportions, the null hypothesis is  $p_1 = p_2$ . There is no need to estimate the individual parameters  $p_1$  and  $p_2$ , but we can estimate their common value with the **pooled sample proportion**. Weighted estimate of  $p_1$  and  $p_2$  is  $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$  and the **standard deviation**  $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p} \cdot \bar{q}}{n_1} + \frac{\bar{p} \cdot \bar{q}}{n_2}}$ .

The test statistic for two proportions with  $H_0: p_1 = p_2$  is 
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p} \cdot \bar{q}}{n_1} + \frac{\bar{p} \cdot \bar{q}}{n_2}}}$$

The following conditions are necessary to use a z-test to test such a difference between the two population proportions.

1. Samples are randomly selected.
2.  $n_1 p_1 \geq 5$ ,  $n_1 q_1 \geq 5$ ,  $n_2 p_2 \geq 5$ , and  $n_2 q_2 \geq 5$ .

3. Samples are independent.

### Guidelines: Two-sample test for the difference between proportions

1. State  $H_0$  and  $H_1$ .
2. Identify  $\alpha$ .
3. Find the test statistic. 
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p} \cdot \hat{q}}{n_1} + \frac{\hat{p} \cdot \hat{q}}{n_2}}}$$
4. Determine the critical value(s). Use table A-2
5. Determine the critical region(s).
6. If  $z$  is in the critical region, reject  $H_0$ . Otherwise, fail to reject  $H_0$ .
7. Interpret the decision in the context of the original claim.

Example 1: In clinical trials of Nasonex, 3774 adult were randomly divided into two groups.

Group 1: (experimental group) received 200 mcg of Nasonex.

Group 2: (control group) received a placebo.

Of the 2103 patients in group 1, 547 reported headaches as a side effect. Of the 1671 patient in group 2, 368 reported headaches as a side effect. Is there sufficient evidence to support the claim that the proportion of Nasonex users that experience headaches is greater than the proportion in the group 2? Use 0.05 as the level of significance.

**Constructing and interpreting confidence intervals** for the difference between two population proportions:

To construct a  $(1-\alpha)100\%$  confidence interval for the difference between two

population proportions, the **margin of error**,  $E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}$  and the

confidence interval for  $p_1 - p_2$  is given by

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

Example 2: In example 1, construct a 90% confidence interval for the difference between the two population proportions.

Example 3: Try it yourself #21, section 9-2

### 9-3 Inferences about two means: Independent Samples

In this section we consider methods for using sample data from two independent samples to test hypotheses made about two population means or to construct confidence interval estimates of the difference between two population means.

**Part I:** Independent samples,  $\sigma_1$  and  $\sigma_2$  unknown and they are assumed not equal.

We can use the following steps to test the claim regarding two population means, provided that

- The samples are obtained using simple random sampling
- Samples are independent
- The populations from which the samples are drawn are normally distributed or the sample sizes are large  $n_1 > 30$  and  $n_2 > 30$ .

Step 1: State the claim mathematically and verbally. The possible pairs of null and alternative hypotheses are:

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Step 2: specify the level of significance.

Step 3: Identify d.f. = smaller of  $n_1 - 1$  and  $n_2 - 1$ .

Step 4: Determine the critical value(s). Use table A-3

Step 5: Determine the critical region(s).

Step 6: Find the test statistic. 
$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Step 7: If  $t$  is in the critical region reject  $H_0$ . Otherwise, fail to reject  $H_0$ .

Step 8: Interpret the decision in the context of the original claim.

Example 4: One study has been conducted to test the effects of marijuana use on mental abilities. Use a 0.01 significance level to test the claim that the population of heavy marijuana users has a lower mean on mental abilities than the light users. Should marijuana use be of concern to college students? The following are the results of memory recall for light and heavy marijuana users.

Items sorted correctly by light marijuana users:

Sample size = 64, sample mean = 53.3, and standard deviation of the sample = 3.6

Items sorted correctly by heavy marijuana users:

Sample size = 65, sample mean = 51.3, and the standard deviation of the sample = 4.5.

Example 5: You want to buy a microwave oven and will choose Model A if its repair costs are lower than Model B's. For 46 Model A ovens, the mean repair cost was \$75 and the standard deviation was \$12.50. For 56 Model B ovens, the mean repair cost was \$80 and the standard deviation was \$20.

At  $\alpha = 0.01$ , would you buy Model A?

**Confidence Intervals:** We can construct confidence interval estimates of the difference between two population means,

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E \quad \text{where} \quad E = t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and}$$

d.f. = smaller of  $n_1 - 1$  and  $n_2 - 1$ .

Example 6: In example 4, construct a 98% confidence interval for the difference between the two population means.

Example 7: As a part of the National Health Survey, data were collected on the weights of men in two different age brackets. For 751 men aged 25-34, the mean is 176 lb and the standard deviation is 35.0 lb. For 1657 men aged 65-74, the mean and standard deviation are 164 lb and 27.0 lb, respectively. a) Test the claim that the younger men come from a population with a mean that is greater than the mean for men in the 65-74 age bracket. Use a 0.01 significance level. b)

Construct a 99% confidence interval for the difference between the means of the men in the two age brackets.

**Part II:** Independent samples,  $\sigma_1$  and  $\sigma_2$  unknown and they are assumed equal.

In this case, the **test statistic is**  $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$  where  $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

and the margin of error:  $E = t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$

**Part III:** Independent samples, and  $\sigma_1$ ,  $\sigma_2$  are known.

In this case, the test statistic is  $z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$  and the margin of error:  $E = z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Example 8: Try it yourself, #17 and 31 from section 9-3.

#### 9-4 Inferences from a matched pairs:

In this section, we will discuss inference on the difference of the two means for dependent samples.

A sampling method is **dependent** when the individuals selected to be in one sample are used to determine the individuals to be in the second sample. For example, if we are conducting a study that compares the IQs of husbands and wives, once a husband is selected to be in the study, his wife is automatically matched with him. Dependent samples are often referred to as **matched-pairs** samples.

**Guidelines:** Testing a claim regarding the mean difference of matched-pairs data, provided that

- The sample is obtained using simple random sampling.
- The sample data are matched pairs.

c) The differences are normally distributed with no outliers or the sample size,  $n$ , is larger than 30.

1. State the claim mathematically and verbally. The possible pairs of null and alternative hypotheses are:

$$H_0: \mu_d = 0$$

$$H_0: \mu_d = 0$$

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

$$H_1: \mu_d < 0$$

$$H_1: \mu_d > 0$$

Where  $\mu_d$  is the mean value of the differences,  $d$ , for the population of all matched pairs.

2. Step 2: specify the level of significance.
3. Step 3: Identify d.f. =  $n-1$ .
4. Step 4: Determine the critical value(s). Use table A-3
5. Step 5: Determine the critical region(s).
6. Step 6: Find the test statistic.  $t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$

$\bar{d}$  is the mean value of the differences,  $d$ , for the paired sample data.

Note:  $\bar{d}$  is equal to the mean of  $(x-y)$  values.

$s_d$  is the standard deviation of the differences,  $d$ .

7. Step 7: If  $t$  is in the critical region reject  $H_0$ . Otherwise, fail to reject  $H_0$ .
8. Step 8: Interpret the decision in the context of the original claim.

Example 9: use a 0.05 significance level to test the claim that when the 13<sup>th</sup> day of a month falls on a Friday, the numbers of hospital admissions from motor vehicle crashed are not affected.

Friday the 6 <sup>th</sup> :	9	6	11	11	3	5
Friday the 13 <sup>th</sup> :	13	12	14	10	4	12

### Constructing and interpreting the confidence intervals about the population mean difference of matched- pair's data:

A  $(1-\alpha)100\%$  confidence interval for  $\mu_d$  is given by:  $\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

The critical value  $t_{\alpha/2}$  is determined using  $n-1$  degrees of freedom.

Example 10: Listed below are sample results of self-reported and measured male heights:

Reported: 68      71   63   70   71   60   65   64   54   63   66  
72

Measured: 67.9      69.9 64.9   68.3   70.3   60.6   64.5   67   55.6   74.2   65  
70.8

- a) Is there sufficient evidence to support the claim that there is a difference between self-reported heights and measured heights of males? Use a 0.05 significance level.
- b) Construct a 99% confidence interval for the difference between the means of the self-reported and measured male heights.

Note that if the confidence interval limits contain 0, indicating that the true value of  $\mu_d$  is not significantly different from 0.

