

CHAPTER 9

9. Comparison of two population Means and Population Proportions

9.1 Inference About Differences Between Two Means

9.1.1 Introduction

Inferences concerning differences between two population parameters are considered as comparative studies. Comparative studies are designed to discover and evaluate differences between groups or between treatments. In comparative study experiments are conducted to collect informative data and conclusions are drawn based on the experimental evidences.

In studies involving comparison of two groups there are two ways of taking the samples and conducting the experiment:

- i. Paired samples and
- ii. Independent samples

Definition:

- i. Two samples are said to be paired if each data point in the first sample is matched and related to a unique data point in the second sample. Pairs of similar individuals (observations) are selected. In an experiment one treatment is applied to one member of each pair and another treatment is applied to the other member. A common application occurs in self pairing where a single individual is measured on two occasions.
- ii. Two samples are independent if data points in one sample are unrelated to data points in the second points. This case arises when we wish to compare two populations and have drawn a sample from each quite independently. The independent samples are used widely when there is no suitable basis for paring.

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

$$H_1: \mu_1 > \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Test statistic: To test the hypotheses we devise a procedure for taking a random sample computing an appropriate test statistic, and then rejecting or failing to reject the null hypothesis H_0 . Part of this procedure is specifying the set of values for the test statistic that leads to rejection of H_0 . This set of values is called the critical region or rejection region for the test.

In order to compare two population means, it is necessary to take two random and independent samples from the two populations. Independence is achieved by making the selection of one sample not influence the selection of the other in any way. We shall use the notations shown in Figure 9.1.

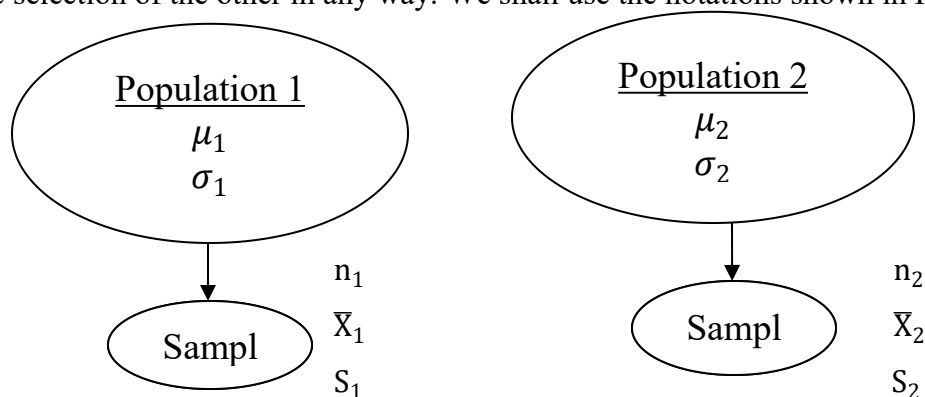


Figure 9.1 Comparing Two Population Means

The hypothesis are about $\mu_1 - \mu_2$, such as

$$H_0: \mu_1 - \mu_2 = D$$

$$H_1: \mu_1 - \mu_2 \neq D$$

For a two-tailed test where D is the hypothesized difference in the means

Often, D is zero.

For a one-tailed test with rejection on the left the hypotheses will be

$$H_0: \mu_1 - \mu_2 = D$$

$$H_1: \mu_1 - \mu_2 < D \text{ and for rejection on the right tail}$$

$$H_0: \mu_1 - \mu_2 = D$$

$$H_1: \mu_1 - \mu_2 > D$$

9.1.2 Sampling Distribution of the Difference Between Two Means

It often becomes important to compare two population means. Knowledge of the sampling distribution of the difference between two means is useful in studies of this type. It is generally assumed that the two populations are normally distributed.

Sampling distribution of $\bar{X}_1 - \bar{X}_2$

Plotting mean sample differences against frequency gives a normal distribution with mean equal to $\mu_1 - \mu_2$ which is the difference between the two population means.

Variance

The variance of the distribution of the sample differences is equal to $\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$. Therefore, the standard error of the differences between two means would be equal to $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Converting to a z score

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

9.1.3 Comparison of Means in Independent Samples

There are many sampling situations in which we will be selecting independent random sample from two populations in order to compare the population means. The statistic used to make such inference will in many cases be the difference in the corresponding sample statistic.

In situation where we wish to make inference about $\mu_1 - \mu_2$, based on independent samples we use the data from the two samples and make a comparison between the population means μ_1 and μ_2 . A logical point estimate of the difference in the population means is the sample difference $\bar{x}_1 - \bar{x}_2$ where \bar{x}_1 is the mean of the sample observations from population 1 and \bar{x}_2 is the mean of the sample observations from population 2.

For confidence interval estimation and hypothesis testing we shall consider the following four cases.

Case 1: small sample and equal variance

The assumption of equal variances is justified when the causes for variations are the same for both populations.

For instance, we may be testing the difference between the average strengths of pins produced from two different raw materials but using the same machinery and process. It is reasonable in this case to assume equal variances.

Assumptions:

1. The samples from the two populations were drawn independently.
2. The population variances/standard deviations are equal.
3. The populations are both normally distributed.

From sampling theory, we note that

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$$

$$V(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

By the Central Limit Theorem, as n_1 and n_2 both increase, $\bar{X}_1 - \bar{X}_2$ will approach the normal distribution.

If random samples of small size are chosen independently from two normal populations with different means but identical variance σ^2 , that is if $x_{11}, x_{12}, \dots, x_{1n_1}$ is a random sample of small size n_1 from $N(\mu_1, \sigma^2)$ and $x_{21}, x_{22}, \dots, x_{2n_2}$ is random sample of small size n_2 from a population which is $N(\mu_2, \sigma^2)$ then we have:

$$\bar{x}_1 = N\left(\mu_1, \frac{\sigma^2}{n_1}\right) \text{ and } \bar{x}_2 = N\left(\mu_2, \frac{\sigma^2}{n_2}\right)$$

Thus the mean of the sampling distribution of $[\bar{x}_1 - \bar{x}_2]$ is normal with mean $E[\bar{x}_1 - \bar{x}_2] = \mu_1 - \mu_2$, and the variance of the sampling distribution of $[\bar{x}_1 - \bar{x}_2]$ is

$$\sigma^2_{(\bar{x}_1 - \bar{x}_2)} = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

Then $Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ has a standard normal distribution.

If the common variance σ^2 is unknown it is estimated by the pooled estimator of the population variance given by:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Decision rule: we shall reject H_0 at α level of significance if :

$$|Z| > Z_{\alpha/2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$Z > Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$Z < -Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence interval: A $100(1-\alpha)\%$ confidence interval for mean difference $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

If σ^2 is unknown $T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Decision rule: we shall reject: we shall reject H_0 at α level of significance if :

$$|T| > T_{\alpha/2, n_1 + n_2 - 2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$T > T_{\alpha, n_1 + n_2 - 2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$T < -T_{\alpha, n_1 + n_2 - 2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence interval:

A 100(1- α)% confidence interval for mean difference $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm T_{\alpha/2, n_1 + n_2 - 2} * S_p \sqrt{\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}}$$

Example 1: Lung destructive index:

Given:

Smokers: $\bar{x}_1 = 17.5$ $n_1 = 16$ $s_1 = 4.4752$

Non-smokers: $\bar{x}_2 = 12.4$ $n_2 = 9$ $s_2 = 4.8492$, $\alpha = 0.05$

We wish to know if we may conclude, at the 95% confidence level, that smokers, in general, have greater lung damage than do non-smokers.

Solution:

Assumptions

- Independent random samples
- Normal distribution of the population
- Population variances are equal and unknown

Hypothesis: $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 > \mu_2$

The appropriate test statistic is:

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

$$\text{Where: } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(15)(4.4711)^2 + (8)(4.8492)^2}{16 + 9 - 2} = \frac{299.86 + 188.12}{23} = 21.2165$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(17.5 - 12.4) - 0}{\sqrt{21.2165 * \frac{1}{16} + \frac{1}{9}}} = \frac{5.1}{1.92} = 2.6563$$

Critical region:

With $\alpha = 0.05$ and $df = 23$, the critical value of t is 1.7139. We reject H_0 if $t > 1.7139$.

Reject H_0 because $2.6563 > 1.7139$

Conclusion: At a significance level of 5% there is sufficient evidence that smokers, in general, have greater lung damage than do non-smokers, i.e., the mean lung damage of smokers is significantly higher from the mean damage of no smokers.

Case 2: Unequal variance but σ_1^2 and σ_2^2 are known

Assumptions:

1. The samples from the two populations were drawn independently.

2. The population variances/standard deviations are NOT equal.
3. The populations are both normally distributed.

When σ_1 and σ_2 are known, we can therefore conduct a z-test if both n_1 and n_2 are more than 30 and the two populations are normal. The z statistic for testing $H_0: \mu_1 = \mu_2$ is given by the formula

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

Decision rule: we shall reject H_0 at α level of significance if :

$$|Z| > Z_{\alpha/2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$Z > Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$Z < -Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence interval: A $100(1-\alpha)\%$ confidence interval for mean difference $\mu_1 - \mu_2$ is given by:

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

Example 1: Serum uric acid levels

Given: $\bar{x}_1 = 4.5$ $n_1 = 12$ $\sigma_1^2 = 1$ $\bar{x}_2 = 3.4$ $n_2 = 15$ $\sigma_2^2 = 1.5$ $\alpha = 0.05$

Is there a difference between the means between individuals with Down's syndrome and normal individuals?

Solution:

Assumptions

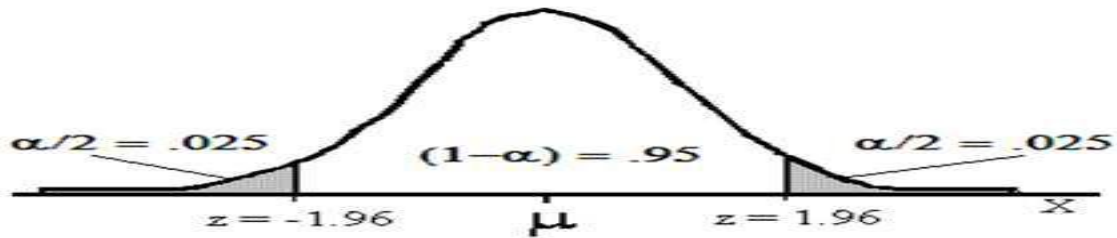
- Two independent random samples
- Each drawn from a normally distributed population with known variance.

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

The appropriate statistic for testing $H_0: \mu_1 = \mu_2$ is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(4.5 - 3.4) - 0}{\sqrt{\frac{1}{12} + \frac{1.5}{15}}} = \frac{1.1}{0.4282} = 2.57$$

Critical region: with $\alpha = 0.05$, the critical values of z are -1.96 and +1.96. We reject H_0 : if $z < -1.96$ or $z > +1.96$.



Reject H_0 because $2.57 > 1.96$.

Conclusion: From these data, it can be concluded that the population means are not equal. At 95% level of confidence there enough evidence that Serum uric acid levels of Down's syndrome and normal individuals are different.

Case 3: Unknown variance but the sample sizes are large

If large samples of size are drawn from two population with unknown variances σ_1^2 and σ_2^2 , an appropriate test statistic for testing $H_0: \mu_1 = \mu_2$ is given by:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{(\bar{X}_1 - \bar{X}_2)}}, \text{ Where } \hat{\sigma}_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$S_1 = \sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2} \text{ and } S_2 = \sqrt{\frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2}$$

Decision rule: we reject: we shall reject H_0 at α level of significane if :

$$|Z| > Z_{\alpha/2} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$Z > Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$Z < -Z_{\alpha} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence interval: A $100(1-\alpha)\%$ confidence interval for mean difference $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} * \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example 1: These data were obtained in a study comparing persons with disabilities with persons without disabilities. A scale known as the Barriers to Health Promotion Activities for Disabled Persons (BHADP) Scale gave the data. We wish to know if we may conclude, at the 99% confidence level, that persons with disabilities score higher than persons without disabilities.

Given

Disable: $\bar{x}_1 = 31.83$	$n_1 = 132$	$s_1^2 = 7.93$	
Non-disable: $\bar{x}_2 = 25.07$	$n_2 = 137$	$s_2^2 = 4.80$	$\alpha = 0.01$

Solution:

Assumptions

- Independent random samples
- large samples
- unknown variance

Hypothesis: $H_0: \mu_1 = \mu_2$ Versus $H_1: \mu_1 > \mu_2$

Test statistic: Because of the large samples, the central limit theorem permits calculation of the z score as opposed to using t. The z score is calculated using the given sample standard deviations. If the assumptions are correct and H_0 is true, the test statistic is approximately normally distributed

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(31.83 - 25.07) - 0}{\sqrt{\frac{(7.93)^2}{132} + \frac{(4.80)^2}{137}}} = \frac{6.76}{0.8029} = 8.42$$

Critical region:

With $\alpha = 0.01$ and a one tail test, the critical value of z is 2.33. We reject H_0 if $z > 2.33$.

Discussion: Reject H_0 because $8.42 > 2.33$.

Conclusion: At 99% level of confidence we conclude that the data support the claim that persons with disabilities score higher than persons without disabilities.

Case 4: Unequal variances with no information about σ_1^2 and σ_2^2 , and small sample size.

There are many situations in which the comparison of means has to be made based on small samples from population with different variances. Among the common situations in which we cannot assume $\sigma_1^2 = \sigma_2^2$ are:

- When the samples come from different types of population as in comparisons made from survey data.
- When computing confidence limits in case in which the population means differ widely the common result that σ changes (though slowly) as μ changes will make us hesitant to assume

$$\sigma_1^2 = \sigma_2^2 .$$

- When one treatment is erratic in its performance sometimes giving high sometimes low responses.

In populations that are markedly skew, the relationship between μ and σ is relatively strong.

An appropriate test statistic for testing $H_0: \mu_1 = \mu_2$ is :

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, t \text{ - does not follow the t - distribution, when } H_0 \text{ is true.}$$

Two different form of the distribution of t have been worked out but both require special tables provided by the authors because both were worked out based on different theoretical backgrounds. An appropriate degrees of freedom that will enable to use the ordinary t-table has been suggested. Then an appropriate degrees of freedom for t is given by:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}$$

Decision rule: we reject: we shall reject H_0 at α level of significance if :

$$|t| > t_{\alpha/2}, v \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$t > t_{\alpha}, v \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$t < -t_{\alpha}, v \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence interval: A $100(1-\alpha)\%$ confidence interval for mean difference $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}, v * \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example 1: We wish to compare the mean gestational age (in weeks) of babies born to women with preeclampsia during pregnancy vs. those who had normal pregnancies. Is the mean gestational age for babies born to preeclamptic mothers is less than the mean gestational age for babies born to mothers with normal pregnancies at 95% confident level?

Data:

Preeclampsia: 38, 32, 42, 30, 38, 35, 32, 38, 39, 29, 29, 32

Normal: 40, 41, 38, 40, 40, 39, 39, 41, 41, 40, 40, 40

Solution:

Preeclampsia: $\bar{x}_1 = 34.5$ $n_1 = 12$ $s_1^2 = 19.36$

Normal: $\bar{x}_2 = 39.92$ $n_2 = 12$ $s_2^2 = 0.81$ $\alpha = 0.05$

Hypothesis: $H_0: \mu_1 = \mu_2$ Versus $H_1: \mu_1 < \mu_2$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(34.5 - 39.92) - 0}{\sqrt{\frac{19.36}{12} + \frac{0.81}{12}}} = \frac{-5.42}{1.297} = -4.177$$

Critical region: $t < -t_{\alpha}, v$, where $df = \text{smaller sample size} = 12$ or

$$\text{Use } v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{19.36}{12} + \frac{0.81}{12}\right)^2}{\frac{\left(\frac{19.36}{12}\right)^2}{12-1} + \frac{\left(\frac{0.81}{12}\right)^2}{12-1}} = \frac{2.8252}{0.237} = 11.9 \approx 12$$

$$t_{\alpha, V} = t_{0.05, 12} = -2.179$$

Reject H_0 because $-4.177 < -2.179$

Conclusion: since $t = -4.177$ is less than -2.179 we reject H_0 at 5% level of significance and conclude that the mean gestational age for babies born to preeclamptic mothers is less than the mean gestational age for babies born to mothers with normal pregnancies.

3.1.4 Comparison of Means in Paired Samples

At times it might be possible to pair the observations in the two samples and take the difference in each pair of observations. Usually this happens when subjects are exposed to a treatment, measurements are taken before and after the treatment, and these measurements are compared to test the effectiveness of the treatment. In effect, this amounts to a single population test where the population is the set of all possible differences in the measurements.

A paired difference test is more efficient than the other tests because it has less chances of Type I and Type II errors for the same sampling effort. The efficiency obtains because when measurements are made on the same subject before and after a treatment, effects of extraneous variables such as age, race and gender on the before/after difference are avoided. Thus the measured difference due to the treatment is more accurate and reliable. Hence, whenever a paired difference test is possible, one should settle for that rather than for other types of tests.

When using dependent samples each observation from population 1 has a one-to-one correspondence with an observation from population 2. One of the most common cases where this arises is when we measure the response on the same subjects before and after treatment.

When two samples are not independent and observations are taken in pairs the paired t-test is applicable. In this case for each data point in one sample there is corresponding data point in the second sample.

Consider n paired sample points $(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})$. Let the mean difference among pairs in the population be denoted by μ_D .

Let $D_i = x_{1i} - x_{2i} \sim N(\mu_1 - \mu_2, \sigma_d^2)$

We desire to test the hypothesis:

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

$H_1: \mu_1 > \mu_2$

$H_1: \mu_1 < \mu_2$

Or equivalently the hypotheses are:

$H_o : \mu_d = 0$

$H_a : \mu_d > 0$ or $H_a : \mu_d < 0$ or $H_a : \mu_d \neq 0$

The mean of difference in the sample is:

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \text{ and the variance of the difference is } S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

Since $D_i \sim N(\mu_1 - \mu_2, \sigma_d^2)$ we have $\bar{D} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_d^2}{n}\right)$

Where σ_d^2 is the variance of the population difference which is seldom known and estimated by the sample variance.

An appropriate test statistic for testing $H_0: \mu_1 = \mu_2$ is given by:

$T = \frac{\bar{D}}{S_d / \sqrt{n}}$ has t-distribution with n-1 under the null hypotheses.

Decision rule: we shall reject H_0 at α level of significance if

$$|T| > T_{\alpha/2, n-1} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 \neq \mu_2$$

$$T > T_{\alpha, n-1} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 > \mu_2$$

$$T < -T_{\alpha, n-1} \text{ for } H_0: \mu_1 = \mu_2 \text{ Versus } H_1: \mu_1 < \mu_2$$

Confidence Interval: A $100(1-\alpha)\%$ confidence interval for mean difference $\mu_1 - \mu_2$ is given

$$\text{by: } \bar{D} \pm T_{\alpha/2, n-1} * \frac{S_d}{\sqrt{n}}.$$

Example 1: Tumor size

Having an accurate measure of tumor size is extremely important because it allows a physician to accurately determine if a tumor is growing, shrinking or remaining constant. The problem is that often the measurements of the tumor size vary from physician to physician. In the past, tumor size was

measured using the linear distance across the tumor, but this was found to be very variable because of the irregular shape of some tumors. A new method called the RECIST criteria traces the outside of the tumor. The RECIST method was believed to give more consistent measures of the volume of the tumor. For a portion of the study, a pair of doctors were shown the same set of tumor pictures. The volume of the tumor was measured by two separate physicians under similar conditions.

Question of interest: Did the measurements from the two physicians significantly differ? If not, then there would be no evidence that the volume measurements change based on physician.

Tumor	1	2	3	4	5	6	7	8	9	10
Dr.1	15.8	22.3	14.5	15.7	26.8	24.0	21.8	23.0	29.3	20.5
Dr.2	17.2	20.3	14.2	18.5	28.0	24.8	20.3	25.4	27.5	19.7

Solution:

We can measure the effect of the treatment in each person by taking the difference $d_i = x_{1i} - x_{2i}$. Instead of having two samples, we can consider our dataset to be one sample of differences

Tumor	1	2	3	4	5	6	7	8	9	10
Dr.1	15.8	22.3	14.5	15.7	26.8	24.0	21.8	23.0	29.3	20.5
Dr.2	17.2	20.3	14.2	18.5	28.0	24.8	20.3	25.4	27.5	19.7
Difference	-1.4	2.0	0.3	-2.8	-1.2	-0.8	1.5	-2.4	1.8	0.8

Volume from Dr. 1: Population mean = μ_1 , Sample mean = \bar{x}_1

Volume from Dr. 2: Population mean = μ_2 , Sample mean = \bar{x}_2

Difference in population mean: $\mu_D = \mu_1 - \mu_2$

Difference in sample mean: $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$

Assuming d_i 's are normally distributed, can use t-distribution with n-1 df where n is the number of difference.

$$t = \frac{\bar{d} - \delta_0}{s_d / \sqrt{n}}$$

Standard deviation of differences $s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$

Test statistic acts just like one sample

1) Null hypothesis: No difference between physicians effect

$$H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0 \text{ versus } H_1: \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$$

2) $\alpha = 0.05$

3) Test statistic

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-0.22}{1.66/\sqrt{10}} = -0.457$$

4) Critical value $t_{0.025,9} = 2.26$ (p-value = 0.53)

5) Decision: we fail to reject null hypothesis

6) Conclusion: there is no evidence of a difference in tumor volume measurement based on physician at 5% level of significance.

Confidence interval for paired t-test constructed in the same way as one-sample t-test

$$\left(\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}}, \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}}\right)$$

For our example, the confidence interval is (-1.01, 0.54)

Note that the conclusion from the hypothesis test and the confidence interval are the same.

Example 2: A physical education director claims that by taking 800 international units (IU) of vitamin E, a weight lifter can increase his strength. Eight athletes are selected and given a test of strength, using the standard bench press. After two weeks of regular training supplemented with vitamin E, they were tested again. Test the effectiveness of vitamin E regiment at 5% level of significance. Each value in the data that follow represents the maximum number of pounds of the athletes can bench press. (Assume that the variable is approximately normally distributed).

Athletes	1	2	3	4	5	6	7	8
Before (X_1)	210	230	182	205	262	253	219	216
After (X_2)	219	236	179	204	270	250	222	216

Solution:

For vitamin E to be effective, the “before weights” must be less than the “after weights”, i.e., the difference in the population means must be negative in order for the vitamin to effective.

Thus the null and alternative hypotheses to be tested are:

- Hypothesis: $H_0: \mu_1 - \mu_2 = 0$ or $\mu_d = 0$ (Vitamin E does not increase strength or vitamin E is not effective)

$H_1: \mu_1 - \mu_2 < 0$ or $\mu_d < 0$ (Vitamin E increase strength i.e. Athletes gain more weight after they are supplemented with vitamin E or vitamin E is effective)

- Level of significance, $\alpha=0.05$
- Appropriate test Statistics is:

$$t_{cal} = \frac{\bar{d}}{S_{\bar{d}}} = \frac{\bar{d}}{S_d / \sqrt{n}}$$

- Computing Critical Value

A one-tailed test is required and $n=8$. Thus,

$$t_{critical} = t_{\alpha}(n - 1) = t_{0.05}(8 - 1) = t_{0.05}(7) = 1.895$$

We shall reject H_0 at 0.05 level of significance if $t < 1.895$

- Computing Calculating value

Athletes	1	2	3	4	5	6	7	8	Total
Before	210	230	182	205	262	253	219	216	
After	219	236	179	204	270	250	222	216	
$d_i = X_{1i} - X_{2i}$	-9	-6	3	1	-8	3	-3	0	-19
d_i^2	81	36	9	1	64	9	9	0	209

Thus,

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} = \frac{\sum_{i=1}^8 d_i}{8} = \frac{1}{8}(-19) = -2.375, \text{ and}$$

$$s_d = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n d_i^2 - \frac{\left(\sum_{i=1}^n d_i \right)^2}{n} \right)} = \sqrt{\frac{1}{7} \left(209 - \frac{(-19)^2}{8} \right)} \approx 4.84$$

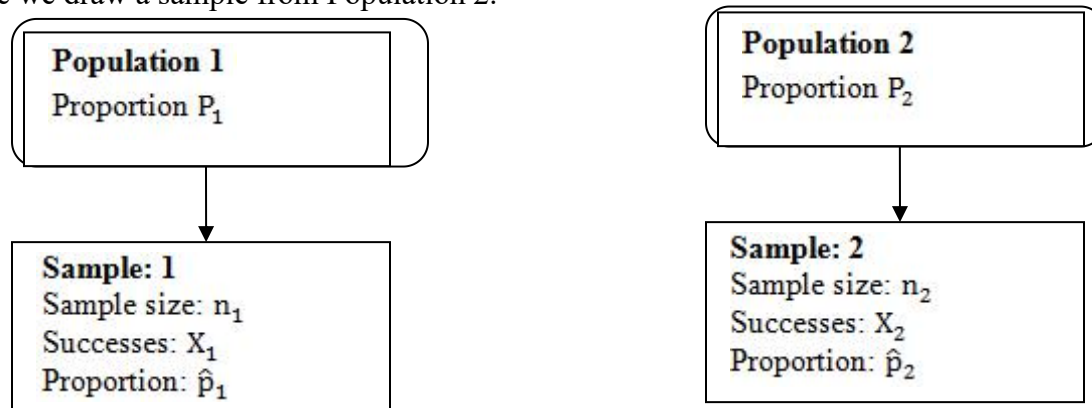
$$\Rightarrow t_{cal} = \frac{\bar{d}}{S_{\bar{d}}} = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{-2.375}{4.84 / \sqrt{8}} = -1.39$$

- Decision Rule: because $t_{cal} > t_{crit}$, we don't reject H_0 at 0.05 level of significance
- Conclusion: There is no sufficient evidence to support the claim vitamin E increases the strength of weight lifters.

9.2 Inference About the Difference Between Two Population Proportions

Introduction

We now consider the case where there are two binomial parameters of interest, say, p_1 and p_2 , and we wish to draw inferences about these proportions. Often you want to analyze differences between two groups in the proportion of items that are in a particular category. The sample statistics needed to analyze these differences are the proportion of occurrences in group 1 and the proportion of occurrences in group 2. With a sufficient sample size in each group, the sampling distribution of the difference between the two proportions approximately follows a normal distribution. Suppose we wish to compare the proportions of two populations that have a specific characteristic, such as the proportion of men who are left-handed compared to the proportion of women who are left-handed. Each population is divided into two groups, the group of elements that have the characteristic of interest (for example, being left handed) and the group of elements that do not. We arbitrarily label one population as Population 1 and the other as Population 2, and we draw a random sample from Population 1 and, without reference to the first sample we draw a sample from Population 2.



Our goal is to use the information in the samples to estimate the difference $P_1 - P_2$ in the two population proportions and to make statistically valid inferences about it.

3.2.2 Sampling Distribution of the Difference Between Two Proportions

We assess the probability associated with a difference in proportions computed from samples drawn from each of these populations.

Sampling distribution of $\hat{P}_1 - \hat{P}_2$.

The sampling distribution of the difference between two sample proportions is constructed in a manner similar to the difference between two means. Independent random samples of size n_1 and n_2 are drawn from two populations of dichotomous variables where the proportions of observations with the character of interest in the two populations are P_1 and P_2 , respectively.

The distribution of the difference between two sample proportions, $\hat{P}_1 - \hat{P}_2$, is approximately normal. For large sample, n_1 and n_2 are large.

The mean for difference $\hat{P}_1 - \hat{P}_2$ is $\mu_{\hat{p}_1 - \hat{p}_2} = P_1 - P_2$

The variance for $\hat{P}_1 - \hat{P}_2$ is $\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}$

The z score for the difference between two proportions is given by:

$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$ has standard normal distribution.

Therefore $\left[(\hat{P}_1 - \hat{P}_2) \sim N\left[\mu = P_1 - P_2, \sigma^2 = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}\right] \right]$ for large sample size

3.2.3 Comparison of Proportions in Independent Samples

Assume we have two binomial populations for which the probability of success in population 1 is P_1 and in population 2 is P_2 . Based on independent samples of size n_1 and n_2 we want to make inferences on the difference between P_1 and P_2 , that is, $(P_1 - P_2)$. The estimate of P_1 is $\hat{P}_1 = X_1/n_1$, where X_1 is the number of successes in sample 1, and likewise the estimate of P_2 is $\hat{P}_2 = X_2/n_2$. Assuming sufficiently large sample sizes, the difference $(\hat{P}_1 - \hat{P}_2)$ is normally distributed with mean $P_1 - P_2$ and variance

$$P_1(1 - P_1)/n_1 + P_2(1 - P_2)/n_2$$

Therefore the appropriate statistic for inferences on $(P_1 - P_2)$ is

$$z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{P_1(1 - P_1)/n_1 + P_2(1 - P_2)/n_2}}$$

Note that the expression for the variance of the difference contains the unknown parameters P_1 and P_2 . In the single-population case, the null hypothesis value for the population parameter p was used in calculating the variance. If the two population proportions are hypothesized to be different, then we substitute \hat{P}_1 and \hat{P}_2 in their places.

If the two population proportions are hypothesized to be equal, then we substitute a pooled proportion \hat{P} in both places. This is analogous to the use of S_1^2 and S_2^2 separately or combining them into S_p^2 in the case of t-test for comparing population means. Letting \hat{P}_1 and \hat{P}_2 be the sample proportions for samples 1 and 2, respectively, the estimate of the common proportion p is a weighted mean of the two-sample proportions,

The pooled proportion is given by the formula

$$\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

where \hat{p} is the average proportion calculated using the formula

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

The test statistic is now computed:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

In construction of a confidence interval for the difference in proportions, we cannot assume a common proportion, hence we use the individual estimates \hat{P}_1 and \hat{P}_2 in the variance estimate. The $(1 - \alpha)$ confidence interval on the difference P_1 and P_2 is:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} * \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$

As in the one-population case the use of the t distribution is not appropriate since the variance is not calculated as a sum of squares divided by degrees of freedom. However, samples must be reasonably large in order to use the normal approximation.

Comparing two population proportions can be done using a z-test if the two samples are sufficiently large. Here a large sample means both np and $n(1 - p)$ are at least 5.

Procedures

- Hypothesis

The null hypothesis is $H_0: P_1 = P_2$ and the alternative hypothesis is one of the following:

$$H_a: P_1 \neq P_2 \text{ Two-tailed}$$

$$H_a: P_1 < P_2 \text{ Left-tailed}$$

$$H_a: P_1 > P_2 \text{ Right-tailed}$$

- Decide on the significance level, α .

- The critical value(s) are

Use Z-table to find the critical value(s)

$$\pm Z_{\frac{\alpha}{2}} \quad \text{For two-tailed}$$

$$- Z_{\alpha} \quad \text{For left-tailed}$$

$$Z_{\alpha} \quad \text{For right-tailed}$$

- Compute the value of the test statistic, $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{P(1-P)}\sqrt{1/n_1 + 1/n_2}}$

- If the value of the test statistic falls in the rejection region, reject H_0 otherwise, do not reject H_0 .

- State the conclusion in words.

The p-value approach compares the p-value for the test statistic with the α level.

Example 1: 200 patients suffering from a certain disease were randomly divided into two groups. Of the first group consisting of 120 patients, who received treatment A, 99 recovered within three days. Out of the other 80, who were treated by treatment B, 62 recovered within 3 days. Can we conclude that treatment A is more effective?

Solution: -Let P_1 = the population proportion of treatment A & P_2 is that of treatment B

Given that $n_1=120$, $n_2= 80$, $x_1= 99$, $x_2= 62$, and $n = n_1+ n_2= 200$. Thus,

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{99}{120} = 0.83, \text{ and}$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{62}{80} = 0.78$$

We need to test: $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$

Level of significance, $\alpha = 0.05$

Test statistics: Since the samples are large, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal.

Under the null hypothesis, the common proportion is estimated by \hat{p}

$$\hat{p} = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2} = \frac{99+62}{200} = 0.81$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.81 = 0.19$$

The test statistics is,

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0.88$$

$Z_\alpha = Z_{0.05} = 1.645$, that is, R: $z > 1.65$ but the calculated value of Z does not lie in the rejection region.

Hence, we fail to reject H_0 and conclude the two treatments are equally effective.

Example 2: A sample of 100 students at the university showed that 43 had taken one or more remedial courses. A sample of 200 students at a junior college showed that 90 had taken one or more remedial college courses. At $\alpha=0.05$, test the claim that there is no difference in the proportion of students who complete remedial course at a university or a junior college.

Solution: Let p_1 = the population proportion of university students who complete remedial courses & p_2 = the population proportion of Junior college students who complete remedial courses. Thus,

The hypothesis to be tested is:

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2$$

Level of significance, $\alpha=0.05$

Test statistics:

Since $n_1=100$ and $n_2=200$ are large, an appropriate test statistics is:

$$\Rightarrow z = \frac{\hat{p}_1 + \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Critical region:

\Rightarrow A two tailed test is required

$\Rightarrow z_{crit} = z_{\frac{\alpha}{2}} = z_{0.025} = 1.96 \Rightarrow \text{Re ject } H_o \text{ if } |z| > 1.96$

Calculated value

Given that $n_1=100$, $n_2=200$, $x_1=45$, $x_2=90$, and $n=n_1+n_2=300$. Thus,

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{45}{100} = 0.43, \text{ and } \hat{p}_2 = \frac{x_2}{n_2} = \frac{90}{200} = 0.45$$

$$\Rightarrow \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{45 + 90}{300} = 0.443$$

$$\text{Thus, } \hat{q} = 1 - \hat{p} = 1 - 0.443 = 0.557$$

$$\text{Hence, } z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -0.33$$

Decision: Since $|Z| < 1.96$, we do not reject H_o at 5% level of significance

Conclusion: There is no difference in proportion of the students who complete remedial course in a university of Junior college.

3.2.4 Comparison of Proportions in Paired Samples

Difference of Two Dependent Proportions

Suppose that a sample of n subjects has been selected to examine the relationship between the presences of a particular attribute at two time points for the same individuals (paired observations). The situation could also be used to examine the relationship between two different attributes for the same individuals.

Attribute at time				
1	2	Number of subjects	Difference(D_i)	Proportion
Present	Present	a	0	a/n
Present	Absent	b	1	b/n

Absent	Present	c	- 1	c/n
Absent	Absent	d	0	d/n
Total		n		1

Let Present = 1 and Absent = 0

Objective: To compare the difference in the proportion of subjects with the attribute at two time points;

$$H_0: P_1 = P_2 \quad \text{Or} \quad H_0: P_1 - P_2 = 0$$

$$H_a: P_1 \neq P_2 \quad H_a: P_1 - P_2 \neq 0$$

The sample difference is $\hat{P}_1 - \hat{P}_2$.

Assumptions

- Pair wise the observations are dependent
- n pairs among themselves are independent
- D_i 's are independent

Then the estimated proportion of subjects with the attribute at time 1 is $p_1 = (a + b)/n$, and the estimated proportion with the attribute at time 2 is $p_2 = (a + c)/n$. The difference between the two estimated proportions is

$$P_1 - P_2 = \frac{a + b}{n} - \frac{a + c}{n} = \frac{b - c}{n}$$

What is the sampling distribution of $\hat{p}_1 - \hat{p}_2$ for large sample size n?

$$\text{Mean} = E[\hat{p}_1 - \hat{p}_2] = E(\bar{D}) = E\left[\frac{1}{n} \sum_{i=1}^n D_i\right] = \frac{1}{n} \sum_{i=1}^n E(D_i), \text{ since identical}$$

$$= \frac{1}{n} \sum_{i=1}^n E(D) = E(D)$$

$$E(\hat{p}_1 - \hat{p}_2) = 0 * P(a) + 1 * P(b) + -1 * P(c) + 0 * P(d) = P(b) - P(c) = p_1 - p_2$$

$$p_1 = \frac{a + b}{n} \text{ and } p_2 = \frac{a + c}{n}$$

$$p_1 - p_2 = \frac{a + b}{n} - \frac{a + c}{n} = \frac{b - c}{n} = P(b) - P(c)$$

Since the two population probabilities are dependent, we cannot use the same approach for estimating the standard error of the difference that we used in the previous section. Instead of showing the steps in the derivation of the formula, we simply present the formula for the estimated standard error.

The appropriate test statistic for testing H_0 is:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - 0}{\sqrt{\frac{P(b) + P(c)}{n}}} \text{ has standard normal distribution.}$$

Let $\hat{P}_1 = \frac{a+b}{n}$ and $\hat{P}_2 = \frac{a+c}{n}$ are small values

$$\hat{P}(b) = \frac{b}{n}, \quad \hat{P}(c) = \frac{c}{n}$$

If $P(b)$ and $P(c)$ are unknown, replaced by $\hat{P}(b)$ and $\hat{P}(c)$ i.e.

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{P(b) + P(c)}{n}}}, \quad P_1 - P_2 = P(b) - P(c)$$

$$\hat{P}_1 - \hat{P}_2 = \hat{P}(b) - \hat{P}(c), \text{ so } Z = \frac{\hat{P}(b) - \hat{P}(c)}{\sqrt{\frac{\hat{P}(b) + \hat{P}(c)}{n}}} = \frac{b - c}{\sqrt{b + c}}$$

With correction for continuity $Z = \frac{b - c - 0.5}{\sqrt{b + c}}$

Confidence Intervals:

The confidence interval for the difference of two dependent proportions $P_1 - P_2$, is then given by

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2} * \frac{1}{n} \sqrt{[b+c] - \frac{[b-c]^2}{n}} \\ = \hat{p}_1 - \hat{p}_2 - Z_{\alpha/2} * \frac{1}{n} \sqrt{[b+c] - \frac{[b-c]^2}{n}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} * \frac{1}{n} \sqrt{[b+c] - \frac{[b-c]^2}{n}} \end{aligned}$$

With correction for continuity $(\hat{P}_1 - \hat{P}_2) \pm Z_{\alpha/2} * \frac{1}{n} \sqrt{[b+c] - \frac{[b-c]^2}{n}} + \frac{1}{2n}$

Example 1: Suppose that 100 students took both calculus and computer tests, and 18 failed in calculus ($\hat{p}_1 = 0.18$) and 10 failed in computer ($\hat{p}_2 = 0.10$). There is an 8 percentage point difference ($\hat{p}_1 - \hat{p}_2 = 0.08$). The confidence interval for the difference of these two failure rates cannot be constructed using the method in the previous subsection because the two rates are dependent.

We need additional information to assess the dependency. Nine students failed both tests ($p_{12}=0.09$), and this reflects the dependency. The dependency between p_1 and p_2 can be seen more clearly when the data are presented in a 2 by 2 table.

Computer			
Calculus	Failed	Passed	Total
Failed	9(a)	9(b)	18
Passed	1(c)	81(d)	82
Total	10	90	100(n)

Solution:

Hypothesis:

$$H_0: p_1 - p_2 = 0 \text{ versus } H_a: p_1 - p_2 \neq 0$$

The marginal totals reflect the two failure rates. The numbers in the diagonal cells (a, d) are concordant pairs of test scores (those who passed or failed both tests), and those in the off-diagonal cells (b, c) are discordant pairs (those who passed one test but failed the other). Important information for comparing the two dependent failure rates is contained in discordant pairs, as the estimated difference of the two proportions and its estimated standard error are dependent on b and c.

$$Z = \frac{b - c - 0.5}{\sqrt{b + c}} = Z = \frac{9 - 1 - 0.5}{\sqrt{9 + 1}} = 2.372$$

Critical value

$$\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96 \Rightarrow -1.96 \leq Z \leq 1.96$$

Since $2.372 > 1.96$ reject the null hypothesis.

Using the standard error equation, we have

$$\text{Estimated SE}(\hat{p}_1 - \hat{p}_2) = \frac{1}{100} \sqrt{[9 + 1] - \frac{[9-1]^2}{100}} = 0.0306$$

Then the 95 percent confidence interval for the difference of these two dependent proportions is $0.08 - 1.96(0.0306) < p_1 - p_2 < 0.08 + 1.96(0.0306)$ or $(0.0200, 0.1400)$.

This interval does not include 0, suggesting that the failure rates of these two tests are significantly different.