In Chapter 9, we have only dealt with a single population. The scenario was that there is a population of interest which is too large for us to observe entirely. Hence population parameters (µ, σ, p) are unknown to us. Instead we gather a random and representative sample from the population, and analyze the sample characteristics to draw conclusions about unknown population characteristics. Ex. we used sample statistics () to test hypotheses formed about the unknown population parameter (µ).

In this chapter, we will be solving similar problems (i.e. running different hypothesis tests). However, we will no longer just be dealing with a single population, but rather make comparisons between two populations.

Consider the following example:

|  |  |
| --- | --- |
| Population 1 | Population 2 |
| Salaries of al males in Bangladesh | Salaries of all females in Bangladesh |
| Mean salary: (unknown) | Mean salary: (unknown) |

We form the hypothesis that the mean salary of males () is greater than mean salary of females (). To test this hypothesis we draw random samples from each of the two populations.



|  |  |
| --- | --- |
| Sample 1: sample of male salaries | Sample 2: sample of female salaries |
| Size: | Size: |
| Mean: | Mean: |
| Standard deviation: | Standard Deviation: |

To run **Difference in Two Population Means** hypothesis tests, we can use two types of samples:

1. Independent Samples
2. Matched Pairs Samples (also called Dependent Samples)

Independent Samples are when Sample 1 and Sample 2 are independent of each other. This means no sample observation in sample 1 influences the selection of observations in sample 2. Ex: If I pick a male of age 25 with a Master’s degree in Sample 1, that does not mean I also have to select a female of age 25 with a Master’s degree in Sample 2.

With reference to the example above, if we use **Independent Samples** then the hypothesis will be written as such:

There are two cases for Independent Samples hypothesis tests:

**Case 1:**

* and are unknown
* and are unknown
* the unknown and are equal to each other, i.e. =
* populations are normally distributed &/or sample sizes of both populations are greater than equal to 30

Use t-distribution to run the Difference in Two Population Means hypothesis test with Independent Samples.

First, convert the difference in sample means to a t test statistic:



in the test statistic is based on the null hypothesis and hence is equal to 0, i.e.

Once the test statistic is computed, apply basic rejection region approach for t-distribution (refer to handout 2).

Use degrees of freedom: *n*1 + *n*2 – 2

**Case 2:**

* and are unknown
* and are unknown
* the unknown and are not equal to each other, i.e.
* populations are normally distributed &/or sample sizes of both populations are greater than equal to 30

Use t-distribution to run the Difference in Two Population Means hypothesis test with Independent Samples.

First, convert the difference in sample means to a t test statistic:



in the test statistic is based on the null hypothesis and hence is equal to 0, i.e.

Once the test statistic is computed, apply basic rejection region approach for t-distribution (refer to handout 2).

Use degrees of freedom:



**Interval Estimate**

Interval Estimate of is a range of values (interval) which contains the true difference in the population means with a probability of . is called the confidence interval, and is usually a large value such as 0.9, 0.95 or 0.99. is called the significance level.

Formula for Interval Estimate is



where t is the critical value that has an area of to its right.



is calculated based on case 1 or 2 mentioned above.