**Notes – Ch10 Inference About Means and Proportions with Two Populations**

Interval estimates and hypothesis tests can be developed for situations involving two populations when the difference between the two population means or the two population proportions is of prime importance. For example, we may want to develop an interval estimate of the difference between the mean starting salary for a population of men and the mean starting salary for a population of women or conduct a hypothesis test to determine whether any difference is present between the proportion of defective parts in a population of parts produced by supplier A and the proportion of defective parts in a population of parts produced by supplier B.

**1. Inferences About the Difference Between Two Population Means: σ1 and σ2 Known**

Letting μ1 denote the mean of population 1 and μ2 denote the mean of population 2, we will focus on inferences about the difference between the means: μ1 - μ2. To make an inference about this difference, we select a simple random sample of n1 units from population 1 and a second simple random sample of n2 units from population 2. The two samples, taken separately and independently, are referred to as **independent simple random samples**.

The difference between the two population means is μ1 - μ2. To estimate μ1 - μ2, we will select a simple random sample of n1 customers from population 1 and a simple random sample of n2 customers from population 2. We then compute the two sample means .

If both populations have a normal distribution, or if the sample sizes are large enough that the central limit theorem enables us to conclude that the sampling distributions of can be approximated by a normal distribution, the sampling distribution of will have a normal distribution with mean given by μ1 - μ2.

We can find the standard error, margin of error and interval estimate.

|  |  |
| --- | --- |
| Point estimator of the difference between the two population means |  |
|  |  |
|  |  |
| Interval estimate of the difference between two population means |  |

**Hypothesis Tests About μ1 - μ2**

Let us consider hypothesis tests about the difference between two population means. Use D0 to denote the hypothesized difference between μ1 and μ2. In many applications, D0 = 0.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Two Tail Test |
| Hypotheses |  |  |  |
| Test Statistics |  |  |  |
| Rejection Rule:  p-value | Reject H0 if | Reject H0 if | Reject H0 if |
| Rejection Rule:  Critical Value | Reject H0 if  z | Reject H0 if  z | Reject H0 if  z |

**2. Inferences About the Difference Between Two Population Means: σ1 and σ2 Unknown**

Letting μ1 denote the mean of population 1 and μ2 denote the mean of population 2, we will focus on inferences about the difference between the means: μ1 - μ2. To make an inference about this difference, we select a simple random sample of n1 units from population 1 and a second simple random sample of n2 units from population 2. The two samples, taken separately and independently, are referred to as **independent simple random samples**.

The difference between the two population means is μ1 - μ2. To estimate μ1 - μ2, we will select a simple random sample of n1 customers from population 1 and a simple random sample of n2 customers from population 2. We then compute the two sample means . We will use the sample standard deviations, s1 and s2, to estimate the unknown population standard deviations. When we use the sample standard deviations, the interval estimation and hypothesis testing procedures will be based on the t distribution rather than the standard normal distribution

We can find the standard error, margin of error and interval estimate.

|  |  |
| --- | --- |
| Point estimator of the difference between the two population means |  |
|  |  |
| Interval estimate of the difference between two population means |  |
| Degree of freedom (df) |  |

We truncate the non-integer degrees of freedom down to provide a larger t-value and a more conservative interval estimate.

**Hypothesis Tests About μ1 - μ2**

Let us consider hypothesis tests about the difference between two population means. Use D0 to denote the hypothesized difference between μ1 and μ2. In many applications, D0 = 0.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Two Tail Test |
| Hypotheses |  |  |  |
| Test Statistics |  |  |  |
| Degree of freedom |  | | |
| Rejection Rule:  p-value | Reject H0 if | Reject H0 if | Reject H0 if |
| Rejection Rule:  Critical Value | Reject H0 if  t | Reject H0 if  t | Reject H0 if |

**3. Inferences About the Difference Between Two Population Means: Matched Samples**

In choosing the sampling procedure that will be used to collect production time data and test the hypotheses, we consider two alternative designs. One is based on independent samples and the other is based on matched samples.

**1. Independent sample design**: A simple random sample of workers is selected and each worker in the sample uses method 1. A second independent simple random sample of workers is selected and each worker in this sample uses method 2.

**2. Matched sample design:** One simple random sample of workers is selected. Each worker first uses one method and then uses the other method. The order of the two methods is assigned randomly to the workers, with some workers performing method 1 first and others performing method 2 first. Each worker provides a pair of data values, one value for method 1 and another value for method 2. In the matched sample design the two production methods are tested under similar conditions (i.e., with the same workers); hence this design often leads to a smaller sampling error than the independent sample design. The primary reason is that in a matched sample design, variation between workers is eliminated because the same workers are used for both production methods.

**Hypothesis Tests About μd**

Let μd = the mean of the difference in values for the population

The d notation is a reminder that the matched sample provides difference data.

The sample mean and sample standard deviation is as follow:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Two Tail Test |
| Hypotheses |  |  |  |
| Test Statistics |  |  |  |
| Degree of freedom | n -1 | | |
| Rejection Rule:  p-value | Reject H0 if | Reject H0 if | Reject H0 if |
| Rejection Rule:  Critical Value | Reject H0 if  t | Reject H0 if  t | Reject H0 if |

**4. Inferences About the Difference Between Two Population Proportions**

Letting p1 denote the proportion for population 1 and p2 denote the proportion for population 2, we next consider inferences about the difference between the two population proportions: p1 - p2. To make an inference about this difference, we will select two independent random samples consisting of n1 units from population 1 and n2 units from population 2.

**Interval Estimation of p1 - p2**

The difference between the two population proportions is given by p1 - p2. The point estimator of p1 - p2 is as follows.

Point Estimator :

Thus, the point estimator of the difference between two population proportions is the difference between the sample proportions of two independent simple random samples. As with other point estimators, the point estimator : has a sampling distribution that reflects the possible values of : if we repeatedly took two independent random samples.

The mean of this sampling distribution = p1 - p2

and the standard error of is as follows:

If the sample sizes are large enough that n1p1, n1(1 - p1), n2 p2, and n2(1 - p2) are all greater than or equal to 5, the sampling distribution of can be approximated by a normal distribution.

An interval estimate is given by a point estimate a margin of error. In the estimation of the difference between two population proportions, an interval estimate will take the following form:

Margin of error

With the sampling distribution of approximated by a normal distribution, we would like to use as the margin of error. The above equation cannot be used directly because the two population proportions, p1 and p2, are unknown. Using the sample proportion to estimate p1 and the sample proportion to estimate p2, the margin of error is as follows.

The general form of an interval estimate of the difference between two population proportions is as follows.

**Hypothesis Tests About p1 - p2**

Let us now consider hypothesis tests about the difference between the proportions of two populations. We focus on tests involving no difference between the two population proportions. In this case, the three forms for a hypothesis test are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Two Tail Test |
| Hypotheses |  |  |  |

When we assume H0 is true as an equality, we have p1 - p2 = 0, which is the same as saying that the population proportions are equal, p1 = p2. We will base the test statistic on the sampling distribution of the point estimator .

The standard error of is as follows:

Under the assumption H0 is true as an equality, the population proportions are equal and

p1 = p2 = p. In this case, becomes

With p unknown, we pool, or combine, the point estimators from the two samples () to obtain a single point estimator of p as follows:

Pooled Estimator of p when p1 = p2 = p

This **pooled estimator of p** is a weighted average of .

Substituting for p in Equation 1, we obtain an estimate of the standard error of .

This estimate of the standard error is used in the test statistic. The general form of the test statistic for hypothesis tests about the difference between two population proportions is the point estimator divided by the estimate of

This test statistic applies to large sample situations where n1p1, n1(1 - p1), n2 p2, and n2(1 - p2) are all greater than or equal to 5.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lower Tail Test | Upper Tail Test | Two Tail Test |
| Hypotheses |  |  |  |
| Test Statistics |  | | |
| Rejection Rule:  p-value | Reject H0 if | Reject H0 if | Reject H0 if |
| Rejection Rule:  Critical Value | Reject H0 if  z | Reject H0 if  z | Reject H0 if  z |