Inference About Two Means: Independent Samples

Gene Quinn

Independent Sampling

We have previously considered **dependent** samples, specifically, samples where subjects were paired in some fashion.

Independent Sampling

We have previously considered **dependent** samples, specifically, samples where subjects were paired in some fashion.

- Same store sales comparisons
- Before and after treatment studies
- Twin studies

Independent Sampling

We now consider the problem of using **independent** samples, that is, samples in which the subjects are chosen independently.

Independent Sampling

We now consider the problem of using **independent** samples, that is, samples in which the subjects are chosen independently.

As with the paired samples, our objective is to test the hypothesis that the population means of two groups are the same:

$$H_0: \mu_1 = \mu_2 \quad \text{versus} \quad H_1: \mu_1 \neq \mu_2$$

or, equivalently,

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

Difference of Two Means -σ Known

The setting for this is the following:

A random sample of size n_1 is taken from population 1, which has mean μ_1 and known standard deviation σ_1.

Difference of Two Means -σ Known

The setting for this is the following:

A random sample of size n_1 is taken from population 1, which has mean μ_1 and known standard deviation σ_1.

A second random sample of size n_2 is taken from population 2, which has mean μ_2 and known standard deviation σ_2.

Difference of Two Means -σ Known

The setting for this is the following:

A random sample of size n_1 is taken from population 1, which has mean μ_1 and known standard deviation σ_1.

A second random sample of size n_2 is taken from population 2, which has mean μ_2 and known standard deviation σ_2.

We assume that either:

The two populations are normally distributed

or

Both samples have at least 30 elements:

$$n_1 \geq 30 \quad \text{and} \quad n_2 \geq 30$$

Difference of Two Means -σ Known

Under these assumptions, the difference of the sample means,

$$D = \overline{x}_1 - \overline{x}_2$$

has a normal distribution with mean:

$$\mu_1 - \mu_2$$

and standard deviation

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Difference of Two Means -σ Known

As usual, we can convert the difference D to a standard normal or Z score by subtracting its mean, and dividing by its standard deviation:

$$Z_D = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Difference of Two Means -σ Known

As usual, we can convert the difference D to a standard normal or Z score by subtracting its mean, and dividing by its standard deviation:

$$Z_D = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The above random variable Z_D has a standard normal distribution.

Difference of Two Means -σ Unknown

In the case where the population standard deviations are not known,

$$D = \overline{x}_1 - \overline{x}_2$$

has a t-distribution *approximately* with mean:

$$\mu_1 - \mu_2$$

and standard deviation

$$\sigma_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$