Testing of Hypothesis: Large Sample Tests

Lesson 2: Test of Hypothesis for Difference of Two Population Means μ_1 and μ_2 (population variances σ_1^2 and σ_2^2 are known)

Let $\overline{x_1}$ and $\overline{x_2}$ are means of two independent samples of sizes n_1 and n_2 (≥ 30). Then the mean of the sampling distribution of difference of means is given by

$$\mu_{\overline{x_1}-\overline{x_2}} = \mu_{\overline{x_1}} - \mu_{\overline{x_2}} = \mu_1 - \mu_2$$

Also the standard deviation of the sampling distribution of difference of means is given by

$$\sigma_{\overline{x_1}-\overline{x_2}} = \sqrt{{\sigma_{\overline{x_1}}}^2 + {\sigma_{\overline{x_2}}}^2} = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

The test statistic $\overline{x_1} - \overline{x_2}$ can then be approximated to a normal variate for large samples and subsequently

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} \sim \mathrm{N}\ (0,1) ------(1)$$

The steps for testing are:

(i) Formulate null hypothesis, H_0 : $\mu_1 = \mu_2$: There is no significant difference between the two population means

(ii) Formulate alternative hypothesis, H_1 : $\mu_1 \neq \mu_2$ (or otherwise) : There is significant difference between the two population means

(iii) Level of significance is : α

(iv) From the two tailed test, the critical region is : $|z| \geq z_{\alpha/2}$

(v) Compute the test statistic z from equation (1)

(vi) Reject H_0 if the computed value of z falls in the critical region, otherwise accept H_0

Note:

(i) If σ_1, σ_2 are unknown, then they can be replaced by sample standard deviations s_1 and s_2 for large sample tests

(ii) If σ^2 is the common variance for the two populations, then equation (1) modifies to

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Problems:

Ex.1. The means of two large samples of sizes 1000 and 2000 are 67.5 and 68 respectively. Test the difference of means of the two populations each with standard deviation 2.5 at 5% level of significance.

Solution:

Here the sample sizes are n_1 = 1000 and n_2 = 2000

The sample means are $\overline{x_1}$ = 67.5 and $\overline{x_2}$ = 68

The common population standard deviation σ = 2.5

(i) Null hypothesis, H_0 : $\mu_1 = \mu_2$, where μ_1 and μ_2 are the two population means

(ii) Alternative hypothesis, H_1 : $\mu_1 \neq \mu_2$

(iii) α = 0.05

(iv) Critical region at 5% level of significance is $|z| \geq 1.96$

(v) $z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(67.5 - 68) - (0)}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.15$

(vi) As computed |z| is more than 1.96, i.e, the value of |z| lies in the critical region, hence null hypothesis H_0 is rejected

Therefore, we conclude that the two population means are not equal.

Ans.

Ex.2. The mean yield of wheat from a district was 210 lbs. with standard deviation 10 lbs. /acre from a sample of 100 plots. In another district, the mean yield was 220 lbs. with standard deviation 12 lbs. /acre from a sample of 150 plots. Assuming that the standard deviation of yield in the entire state was 11 lbs., test whether there is any significant difference between the mean yields of crops in the two districts. (at 1% significant level)

Solution:

Here the sample sizes are n_1 = 100 and n_2 = 150, hence large sample test is to be applied

The sample means are $\overline{x_1}$ = 210 lbs. and $\overline{x_2}$ = 220 lbs.

The sample standard deviations are s_1= 10 lbs. /acre and s_2= 12 lbs. /acre [*No role in calculation as σ is given]

The common population standard deviation σ = 11 lbs.

(i) Null hypothesis, $H_0 : \mu_1 = \mu_2$, where μ_1 and μ_2 are the mean yields of crops in the two districts

(ii) Alternative hypothesis, $H_1 : \mu_1 \neq \mu_2$

(iii) $\alpha = 0.01$

(iv) Critical region at 1% level of significance is $|z| \geq 2.58$

(v) $$z = \frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(210-220)-(0)}{11\sqrt{\frac{1}{100}+\frac{1}{150}}} = -7.04$$

(vi) As computed |z| is more than 2.58, i.e, it lies in the critical region, so null hypothesis H_0 is rejected

Hence, we conclude that there is a significant difference between the mean yields of crops in the two districts.

Ans.

Ex.3. In a certain factory, there are two different processes of manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 gm. with a standard deviation of 12 gm. The corresponding figures in a sample of 400 items from the other process are 124 and 14. Is this difference significant?

Solution:

Here the sample sizes are n_1 = 250 and n_2 = 400, hence large sample test is to be applied

The sample means are $\overline{x_1}$ =120 gm. and $\overline{x_2}$ = 124 gm.

The sample standard deviations are s_1= 12 gm. and s_2= 14 gm.

Since the population standard deviations are not known, they can be replaced by sample standard deviations

(i) Null hypothesis, H_0 : $\mu_1 = \mu_2$, where μ_1 and μ_2 are the average weights of items produced from the two processes

(ii) Alternative hypothesis, H_1 : $\mu_1 \neq \mu_2$

(iii) $\alpha = 0.05$ [* since nothing is mentioned, we take 5% significance level]

(iv) Critical region at 5% level of significance is $|z| \geq 1.96$

(v) $$z = \frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}} = \frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}} = \frac{(120-124)-(0)}{\sqrt{\frac{12^2}{250}+\frac{14^2}{400}}} = -3.9$$

(vi) As computed |z| is more than 1.96, i.e, it is in the critical region, so null hypothesis H_0 is rejected

Hence, we conclude that there is a significant difference between the average weights of items produced from the two processes.

Ans.