Z – test

Test of significance for Difference of Means:

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2 and $\bar{x}_2$ be the mean of an independent random sample of size n_2 from another population with mean μ_2 and variance σ_2^2. Then since sample sizes are large,

$$\bar{x}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Also $\bar{x}_1 - \bar{x}_2$, being the difference of two independent normal variates is also a normal variate. Then Z, (standard normal variate) corresponding to $\bar{x}_1 - \bar{x}_2$ is given by

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)}$$

S.E stands for standard error

$$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2 = 0$$

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

Thus under the null hypothesis $H_0: \mu_1 = \mu_2$ the test statistics becomes (for large samples)

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Q) The means of two single large samples of 1000, and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches? Test at 5% level of significance.

Sol :- Given, $n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$ inches, $\bar{x}_2 = 68.0$ inches

Null hypothesis $H_0 : \mu_1 = \mu_2$ and $\sigma = 2.5$ inches, i.e. the samples have been drawn from the same population of standard deviation 2.5 inches.

Alternative hypothesis : $H_1 : \mu_1 \neq \mu_2$ (Two tailed)

Test statistics, Under H_0, the test statistics is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{67.5 - 68.0}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}}$$

$$= \frac{-0.5}{2.5 \times \sqrt{\frac{1}{1000} + \frac{1}{2000}}} = \frac{-0.5}{2.5 \times 0.0387} = -5.1$$

$|Z| > 3.$

The calculated Z is highly significant and we reject the null hypothesis and conclude that samples are certainly not from the same population with standard deviation 2.5 inches.

Remark ① In the above problem

$\sigma_1^2 = \sigma_2^2 = \sigma^2$, i.e. if the samples have been drawn from the populations with common S.D σ then under $H_0 : \mu_1 = \mu_2$

$$Z = \frac{x_1 - x_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

② If $\sigma_1^2 \neq \sigma_2^2$ and σ_1 and σ_2 are unknown, then they are estimated from sample values. This result in some error, which is practically immaterial if sample are large.

These estimates for large samples are given by $\hat{\sigma}_1^2 = S_1^2 \approx s_1^2$ and $\hat{\sigma}_2^2 = S_2^2 = s_2^2$ (for samples are large)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Ex(2) In a survey of buying habits, 400 women shoppers are chosen at random in super market 'A' located in a certain section of the city. Their average weekly food expenditure is Rs 250 with a standard deviation of Rs 40. For 400 women shoppers chosen at random in super market 'B' in another section of the city, the average weekly food expenditure is Rs 220 with a standard deviation of Rs 55. Test at 1% level of significance whether the average weekly food expenditure of two population shoppers are equal.

Sol: $n_1 = 400$, $\bar{x}_1 = 250$, $s_1 = 40$

$n_2 = 400$, $\bar{x}_2 = 220$, $s_2 = 55$

Null hypothesis $H_0: \mu_1 = \mu_2$ i.e, the average weekly food expenditures of the two populations of shoppers are equal.

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$ (Two tailed)

$$\cancel{z} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Since σ_1^2 and σ_2^2 are not known, we can take $\sigma_1^2 = s_1^2$ and $\sigma_2^2 = s_2^2$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{250 - 220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}}$$

$$= \frac{30 \times 20}{\sqrt{40^2 + 55^2}} = \underline{8.82}$$

Calculated $|Z| > 2.58$

∴ Null hypothesis is rejected
So average expenditure of two populations of shoppers in market A and market B differ significantly.

③ The average hourly wages of a sample of 150 workers in plant 'A' was Rs 2.56 with a standard deviation of Rs 1.08. The average hourly wage of a sample of 200 workers in plant B was Rs 2.87 with a standard deviation of Rs 1.28. Can an applicant safely assume that hourly wages paid by plant 'B' are higher than those paid by plant 'A'?

Sol:- $n_1 = 150$, $\bar{x}_1 = 2.56$, $s_1 = 1.08$
$n_2 = 200$, $\bar{x}_2 = 2.87$, $s_2 = 1.28$

Null hypothesis : $H_0 : \mu_1 = \mu_2$, there is no significant difference between the mean wages of workers in plant A and plant B.

Alternative hypothesis H_1 ; $\mu_1 < \mu_2$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.56 - 2.87}{\sqrt{\frac{(1.08)^2}{150} + \frac{(1.28)^2}{200}}}$$

$$= \frac{-0.31}{\sqrt{0.016}} = -2.46$$

Here calculated $Z < -1.645$ (Critical value of Z for left tailed at 5% level of sig)

So null hypothesis rejected

We conclude that average hourly wage at plant B is certainly higher.