Test of Significance: Large Samples (Z-test)

- Test of significance for single mean
- Test of significance for difference of means of two large samples
- Test of significance for a single proportion
- Test of significance for difference of proportions

Test of significance for single mean

Let $x_1, x_2, \dots, x_n$ be a random sample of size n, drawn from a large population with mean μ and variance σ^2.
Let $\bar{x}$ denote the mean of the sample and s^2 denote the variance of the sample.
We know that $\bar{x} \sim N(\mu, \sigma^2/n)$. The standard normal variate corresponding to $\bar{x}$ is
$Z = \frac{\bar{x}-\mu}{S.E.(\bar{x})}$, where S.E.($\bar{x}$) $=\sigma/\sqrt{n}$.
We set up the null hypothesis that there is no difference between the sample mean and the population mean. The test statistic is

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ If σ is known.

$Z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ If σ is not known. Here, s is the standard deviation of the sample.

Problem 1: The heights of college students in a city are normally distributed with S.D. 6 cms. A sample of 100 students have mean height 158 cms. Test the hypothesis that the mean height of college students in the college is 160 cms.

Solution:

We have $\bar{x} = 158$ (mean of the sample),

$\mu = 160$ (mean of the population), $\sigma = 6, n = 100.$

Level of significance: 5%

$H_0: \mu = 160$, i.e., difference is not significant.

$H_1: \mu \neq 160$

We apply the two tailed test.

Test statistic is $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{158-160}{6/\sqrt{100}} = -3.33.$

$\therefore |Z| = 3.333$

Table value of Z at 5% level of significance $= 1.96$. Since calculated value of Z at 5% level of significance is greater than the table value of Z, we reject H_0 at 5% level of significance.

Problem 2: A sample of 400 items is taken from a population whose standard deviation is 10. The mean of the sample is 40. Test whether the sample has come from the population with mean 38. Also calculate 95% confidence interval for the population mean.

Solution:

$H_0: \mu = 38$

$H_1: \mu \neq 38$

Level of significance: 5%

We apply the two tailed test.

Test statistic is $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{40-38}{10/\sqrt{400}} = 4.$

$\therefore |Z| = 4$

Table value of Z at 5% level of significance $=$ 1.96. Since calculated value of Z at 5% level of significance is greater than the table value of Z, we reject H_0 at 5% level of significance. 95% confidence interval for the population mean is given by

$$\bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}} = 40 \pm \left(1.96 \times \frac{10}{\sqrt{400}}\right) = [39.02,40.98]$$

Problem 3:The mean of a certain normally distributed production process is known to be 50 with a standard deviation of 2.5. The production manager may welcome any change in the mean value towards the higher side but would like to safeguard against decreasing values of mean. He takes a sample of 12 items that gives a mean value of 46.5. What inference should the manager take for the population process on the basis of sample results. Use 5% level of significance for the purpose.

Solution:

$H_0: \mu = 50$

$H_1: \mu < 50$

Level of significance: $\alpha = 0.05$

Test statistic is $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{46.5-50}{2.5/\sqrt{12}} = -4.854.$

$\therefore |Z| = 4.854$

Table value of Z at 5% level of significance $= -1.645$. Since calculated value of Z at 5% level of significance is greater than the table value of Z, we reject H_0 at 5% level of significance.

Problem 4: A sample of 900 members has a mean of 3.4 cms and a S.D. 2.61 cms. Is the sample from a large population of mean 3.25cm?

Solution: Given : $n = 900, \mu = 3.25, \bar{x} = 3.4 \text{ cm}, \sigma = 2.61, s = 2.61$

Null hypothesis : $H_o: \mu = 3.25$

Alternative hypothesis: $H_1: \mu \neq 3.25$

Test statistic (C.V) : $z = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}} = 1.724$

$$\mathrm{Z} = 1.724$$

Tabulated Value (T.V) : Z at 5% level of significance is 1.96

Conclusion : $\mathrm{C.V} < \mathrm{T.V}$

we accept the null hypothesis H_o.

Therefore, the samples has been drawn from the population mean $\mu = 3.25$

Test of significance for difference of means of two large samples

Let $\overline{x_1}$ be the mean of an independent random sample of size n_1 from a population with mean μ_1 and variance σ_1^2. Again, let $\overline{x_2}$ be the mean of an independent random sample of size n_2 from a population with mean μ_2 and variance σ_2^2. Here, n_1 and n_2 are large. Clearly,

$$\overline{x_1} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \overline{x_2} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right).$$

Hence, under the null hypothesis $H_0: \mu 1 - \mu 2 = \delta_0$, the test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2} - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If $\sigma_1 = \sigma_2 = \sigma$, then the test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2} - \delta_0}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If σ_1 and σ_2 are not known use the samples' standard deviation, then the test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2} - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

If $\sigma_1 = \sigma_2 = \sigma$ and σ is not known, we compute σ^2 by using the formula

$$\sigma^2 = \frac{n_1^2 s_1^2 + n_2^2 s_2^2}{n_1 + n_2}.$$

In this case, the test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2} - \delta_0}{\sqrt{\frac{n_1^2 s_1^2 + n_2^2 s_2^2}{n_1 + n_2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Problem 1: Random samples drawn from two places gave the following data relating to the heights of children

	Place A	Place B
Mean height	68.50	68.58
Standard deviation	2.5	3.0
Sample size	1200	1500

Test at 5% level that the mean height is the same for the children at two places.

Solution: $\overline{x_1} = 68.50$, $\overline{x_2} = 68.58$,
$n_1 = 1200$, $n_2 = 1500$
$\sigma_1 = 2.5$, $\sigma_2 = 3.0$
Level of significance$= 0.05$
Null hypothesis: $H_0 : \mu_1 = \mu_2$
Alternative hypothesis: $H_1 : \mu_1 \neq \mu_2$

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = -0.756.$$

$\therefore |z| = 0.756 < z_\alpha = 1.96$. Hence, null hypothesis is accepted.

Problem 2: The means of 2 large samples 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 S.D. 2.5 inches.

Solution : Sample sizes $n_1 = 1000, n_2 = 2000.$

Sample mean $\bar{x}_1 = 67.5$ inches, $\bar{x}_2 = 68$ inches

Population S.D $\sigma = 2.5$ inches

Null Hypothesis : $H_o: \mu_1 = \mu_2$

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$

Test statistic (C.V) : $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.16 \; |z| = 5.16$

Tabulated value (T.V): z at 5% level of significance is 1.96

Conclusion : C.V > T.V

We reject the null hypothesis H_o.

Therefore, the samples are not drawn from the same population.

Problem: Two independent random samples of sizes 10 and 12 from $\mathrm{N}(\mu_1, \sigma_1^2 = 4)$ and $\mathrm{N}(\mu_2, \sigma_2^2 = 9)$ give the sample means 20 and 22 respectively. Test the hypothesis that $\mu_1 - \mu_2 = 5$ against the alternative that $\mu_1 - \mu_2 < 5$ at 10% level.

Solution: The appropriate test-statistic is

$$\tau = \frac{(\bar{x}_1 - \bar{x}_2) - 5}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

which follows, under H_0, standard normal distribution.
From the given data we have,

$$n_1 = 10, n_2 = 12, \sigma_1^2 = 4, \sigma_2^2 = 9, \bar{x}_1 = 20 \text{ and } \bar{x}_2 = 22$$

Hence

$$\tau = \frac{(20 - 22) - 5}{\sqrt{\frac{4}{10} + \frac{9}{12}}} = -\frac{7}{\sqrt{1.15}} = -6.53$$

Here τ(=-6.53)<-1.28. So, reject H_0 at 10% LOS and conclude that $\mu_1 - \mu_2 < 5$.