19/06/2024

Unit - V : Testing of Hypothesis

Statistical Hypothesis

In many circumstances to arrive at decisions about the population on the basis of sample info, we make assumptions about the population parameters. Such assumption is called a Statistical Hypothesis which may be true or may be false.

The procedure which enables us to decide on the basis of sample results whether a hypothesis is true or not is called test of hypothesis.

Procedure in testing of hypothesis

① Null Hypothesis:-

- we first set up a hypothesis, a definite statement about a population parameter. such a hypothesis is usually a hypothesis of no difference.
- It is denoted by H_0
- It is a hypothesis which is tested for possible rejection under the assumption that it is true.

② Alternative Hypothesis:-

- Any hypothesis which is complimented to Null Hypothesis is called Alternative Hypothesis
- It is denoted by H_1
- It is set in such a way that the rejection of Null hypothesis implies that acceptance of alternative hypothesis.

$H_1: \mu \neq \mu_0$ [Two tailed test]

$H_1: \mu > \mu_0$ [Right tailed]

$H_1: \mu < \mu_0$ [left tailed]

③ Test Statistic:

- It is a statistic based on appropriate probability distribution.
- It is used to test whether null hypothesis should be accepted or rejected
- For Z-distribution

  under the normal curve for $n \geq 30$

  test statistic

$$Z = \frac{t - E(t)}{S \cdot E(t)}$$

Errors in Hypothesis Testing:

① Type I Error:

It is error of rejecting null hypothesis when it is true

$P(\text{type I error}) = P(\text{Reject } H_0 \text{ when it is true}) = \alpha$

② Type II error:

It is error of accepting null hypothesis when it is false

$P(\text{type II error}) = P(\text{Accept } H_0 \text{ when it is false}) = \beta$

Here, $\alpha \Rightarrow$ ~~pons~~ producer risk

$\beta \Rightarrow$ consumer risk

④ Level of Significance:

- It is the maximum probability of type I error
- By default $\alpha = 5\%$

  This means there is a probability of making 5 errors out of 100.

Critical values:-

	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
two tailed	$Z_\alpha = 2.58$	$Z_\alpha = 1.96$	$Z_\alpha = 1.645$
Right tailed	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left tailed	$Z_\alpha = -2.33$	$Z_\alpha = -1.28$	$Z_\alpha = -1.28$

⑤ Conclusions:-

If $|Z_{cal}| < Z_\alpha$ then Null Hypothesis is accepted i.e Alternative Hypothesis is rejected

If $|Z_{cal}| > Z_\alpha$ then Null Hypothesis is rejected i.e Alternative Hypothesis is accepted

Test of significance of a single mean:- [Large sample i.e $n \geq 30$]

Null Hypothesis H_0 : $\mu = \mu_0$

A.H H_1 : $\mu \neq \mu_0$
$\mu > \mu_0$
$\mu < \mu_0$

L.O.S : $\alpha = 5\%$

Test Statistic: $Z = \dfrac{\bar{x} - E(\bar{x})}{S.E(\bar{x})} \Rightarrow Z_{cal} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

conclusion: If $|Z_{cal}| < Z_\alpha$ then H_0 is accepted otherwise H_1 is accepted

Q) A sample of 400 items is taken from a population whose S.D is 10 the mean of sample is 40. Test whether the sample has come from a population with mean 38 at 5% ~~error~~ level of significance. Also calculate 95% confidence interval

A) $n = 400$

$\sigma = 10$

$\bar{x} = 40$

N.H $\quad H_0$: The sample is drawn from the population of mean 38 i.e $\mu = 38$

A.H $\quad H_1$: The sample is not drawn from the population of mean 38 i.e $\mu \neq 38$ [two tailed]

L.OS :- $\alpha = 5\% \longrightarrow Z_\alpha = 1.96$

T.S :- $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{40 - 38}{10/\sqrt{400}} = 4$

Conclusion :- $|Z_{cal}| > Z_\alpha$ (i.e $4 > 1.96$)

Null Hypothesis is rejected ie Alternative Hypothesis is accepted

that means the sample is not drawn from the given population

95% confidence interval for population mean

$\left(\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 40 \pm 1.96 \left(\frac{10}{\sqrt{400}} \right) = (39.02, 40.98)$

Q) In a random sample of 60 workes the average can be taken by them to get work done is 33.8 minutes, with S.D of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of A.H $\mu > 32.6$ minutes at 1% level of significance.

A) $n = 60$

$\bar{x} = 33.8$

$\sigma = 6.1$

N.H $H_0 : \mu = 32.6$

A.H $H_1 : \mu > 32.6$ [Right tailed]

$\alpha = 1\% \longrightarrow Z_\alpha = 2.33$

$$Z_{cal} = \frac{33.8 - 32.6}{6.1/\sqrt{60}}$$

$$= 1.523$$

Since $Z_{cal} < Z_\alpha$, N.H is accepted

- Test of Significance of difference of Means [Large sample]

N.H $H_0 : \mu_1 = \mu_2$

A.H $H_1 : \mu_1 \neq \mu_2$ [two tailed]

$\mu_1 < \mu_2$ [Left tailed]

$\mu_1 > \mu_2$ [right tailed]

L.O.S : $\alpha = 5\% \rightarrow Z_\alpha = 1.96$

$$T.S \Rightarrow Z = \frac{\bar{x}_1 - \bar{x}_2 - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If two samples are drawn from same population of S.D σ

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

If population S.D's are not ~~sam~~ given

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

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

where s_1^2 – variance of first sample
s_2^2 – variance of second sample

Q) Two types of new cars produced in USA and tested for petrol mileage. One sample is consisting of 42 cars average 15 kmpl while the other sample is consists of 80 cars – 11.5 kmpl with population variances as $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$ test whethere there is any significant difference in the petrol consumption of the two types of car at 1% level of significance

A) N.H H_0: There is no significant difference in the consumption of petrol of two models

i.e $\mu_1 = \mu_2$

A.H H_1: $\mu_1 \neq \mu_2$

L.O.S: $\alpha = 1\% \rightarrow Z_\alpha = 2.58$

T.S ⇒ $Z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{15 - 11.5}{\sqrt{\dfrac{2.0}{42} + \dfrac{1.5}{80}}}$

$n_1 = 42$
$n_2 = 80$
$\overline{x_1} = 15$
$\overline{x_2} = 11.5$

$= 13.587$

∴ ⇒ H_1 is accepted