Testing Hypothesis

1) Population :
Population is set of all observations under study.

2) Sampling :
It is the process of selecting samples from the population.

3) Random sampling :
The selection of samples at random avoiding the possibility of biasness is called random sampling.

4) statistical inference :
The generalizes of inferences from the sample to the population is called statistical inference.

5) Parameters :
The ~~statical~~ statistical measures of population such as mean (μ), varience(σ^2), are called the parameters of population.
similarly, statistical measures computed from sample such as mean ($\overline{x}$), varience ~~($\overline{a}^2$)~~ (s^2) are called statistics.

6) sampling distribution :
Let all the possible samples of size n are drawn from the population at random. Then we compute mean ($\overline{x}$) for each sample.

All the mean~~s~~ will not be the same.
If these different mean are grouped according to their frequencies, then frequency distribution which obtained is called sampling distribution of mean.

similarly, we can form sampling distribution of varience.

A sample having size $n \geq 30$ is called a large sample, otherwise a small sample. ~~It t~~ If the sample is large then sampling distribution of statistics approaches a normal distribution.

~~Te~~
Test of significance–

This test enables us to side on the basis of sample result whether the difference between observed

~~a~~ or difference between 2 independant sample statistics is significant or might be attributed due to chance or fluctuations of ~~sample~~ sampling.

Standard Error–

The standard defination of the sampling distribution of a statistics is called standard error of statistics.

$A \cup B = \{S\} \Rightarrow$ Exhaustive.

The reciprocal of standard error is called precision.

Statistics	Error
1) Difference betn $\bar{x}$ & μ	1) $\sigma/\sqrt{n}$
2) Difference betn sample standard deviation (s) & population standard deviation (σ)	2) $\sigma/\sqrt{2n}$
3) Difference betn sample proportion (p) & population proportion (P)	3) $\sqrt{PQ/n}$ $P+Q=1$ $Q=1-P$.
4) Difference betn two sample mean. $(\bar{x}_1-\bar{x}_2)$	4) $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
5) Diffn betn two sample standard deviation (s_1-s_2)	5) $\sqrt{\frac{\sigma_1^2}{2n_1}+\frac{\sigma_2^2}{2n_2}}$
6) Diffn betn two sample proportions (p_1-p_2)	6) $\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

Null hypothesis :

For applying the test of significence we set a hypothesis which is tested for possible rejection ~~assumptions~~ under assumption of it is true. such hypothesis is called null hypothesis.

The hypothesis complimentary to hypothesis called null hypothesis ~~is~~ it is ~~called~~ denoted by h_1.

suppose, we want to test the null hypothesis of assumed that null hypothesis of μ_0. H_0 is $\mu = \mu_0$

The 3 possible alternative hypothesis will be

1) $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$) is called 2 tailed hypot

2) $H_1 : \mu > \mu_0 \Rightarrow$ Right tailed

3) $H_1 : \mu < \mu_0 \Rightarrow$ left tailed.

Critical region :

A region 'R' corresponding to a sample statistics 't' in the sample space which amoungs to rejection of hypothesis is called critical region

The region complemantry to critical region is called acceptance region.

Label of significance –

The probability of the value of variant falling in critical region is called label of significance.

If p is the value of ~~statics~~ statistics obtained using random sample of size n & r is the critical region then the probability α that a random sample of the statistics 't' belongs to the critical region is the level of significance & is given by $P(t / H_0) = \alpha$

The level of significance is always fixed in advance before studying the charactieristics of random sample. It is usually expressed

Total area of critical region is written as α% ~~level~~ label of significance. Generally, 2 popular values of ~~le~~ label of significance is 1% & 5%.

Single tail & two-tailed test :

A test of any statistical hypothesis in which the alternative hypothesis is single tailed is called single tailed test.

Otherwise, it is called as two-tailed test.

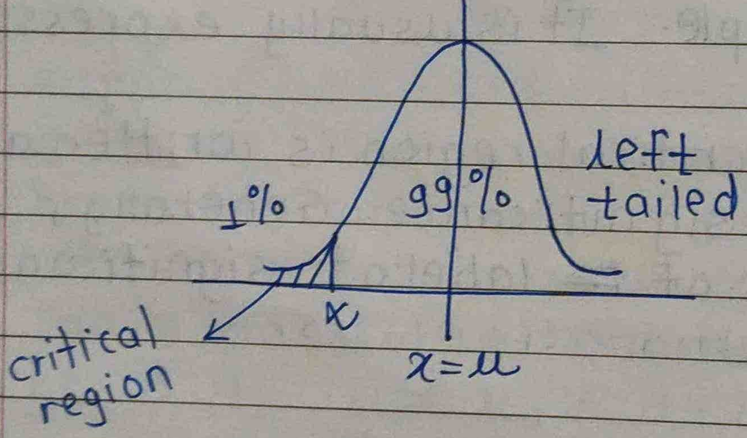
Critical value :

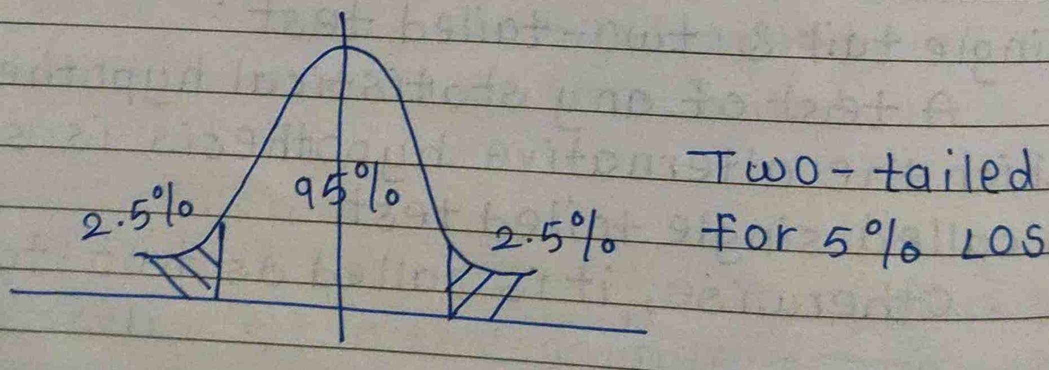
In case of large samples, if any statistics 't' ~~is~~ & $E(t)$ is the corresponding population mean then the variable $z = \dfrac{t - E(t)}{SE(t)}$ is normally distributed with mean zero & varience unity. The value of test statistics z which separates critical region & accepled region is called critical value or significant of z. value dentoned by $z(\alpha)$. & α is ~~level~~ lebal of significance.

The critical value depends on

1) The prescribed lebal of significance.
2) Alternative hypothesis whether it is two tailed / single tailed.
3) The critical value $z(\alpha)$ of the test statistics for 2 tailed test is given by

$$P(|Z| > Z_\alpha) = \alpha.$$





~~Two tailed test:~~

Test	Critical value	Level of significance 1%	5%	10%
Two-tailed test	$\lvert z_\alpha \rvert$	2.58	1.96	1.645
Right-tailed test	z_α	2.33	1.645	1.28
left-tailed test	$-z_\alpha$	-2.33	-1.645	-1.28

Confidence limit :

——— " ——— & end points of confidence variable i.e. the interval in which population parameter is suppose to lies.

Confidence level :

The probability i.e. associated with confidence interval is called confidence level is return on $1-\alpha$

Consider, the sampling distribution of statistics t is normal ~~wh~~ with mean μ & standard deviation σ.

The sample statictics can be expected to lie in the interval $(\mu - 1.96\sigma, \mu + 1.96\sigma)$ for 95% cases

Errors in testing hypothesis :

The main aim of sampling is to draw valid inferences about the population paramet on the basis of sample ~~resut~~ reasons. Because of these reasons we are able to commite following types of errors :

1) Type-1 Error –

We reject H_0 when it is true. If we write probability that reject H_0 for given H_0

$P(\text{Reject } H_0/H_1)=\alpha$ then α is called size of type-1 error which is also referred as producer's risk.

2) Type-2 Error –

We accept H_0, when it is not true i.e. accept H_0 when H_1 is true. If we write ~~probability~~ $P(\text{accept } H_0/H_1)=\beta$ then β is called as size of type-2 error & is refeired as consumer's risk.

Steps for test hypothesis :

1) Define null hypothesis H_0.
2) —"— alternative hypothesis H_1 so, as to decide the test to be use, two-tailed / single-tailed.
3) Depending upon the problem fix the appropriate level of significance.
4) obtain the value of Z_α at the level of significance.
5) Compute the test statistics $Z=\dfrac{t-E(t)}{SE(t)}$ under the null hypothesis.

} Any 1 is t.

P - Population
p - sample.

6) compare z with z_α for corresponding level of significance if $|z| > z_\alpha$ the H_0 is rejected & H_1 is accepted. that means the difference $|t - E(t)|$ is significant. or if $|z| < z_\alpha$ then we accepte H_0 & reject H_1 for the level of significance. This implies the difference ~~$E(t)$~~ $|t - E(t)|$ is not significant.

Test for difference bet^n sample proportion or population proportion.

Let P be population proportion & p be sample proportion.
Let sample of size n be drawn from the population.
If x is no. of successes n independent trial with constant probability p for success for each trial then $X \sim \beta(n, p)$

$X \sim \beta(n, p)$ $\quad\quad Q = 1 - p$

$E(X) = nP$

$V(x) = nPQ$

$P = \frac{X}{n}$

then $E(P) = E\left(\frac{X}{n}\right)$

$= \frac{1}{n} E(X)$

$= \frac{1}{n} (np)$

$= p.$

$$v(x) = v\left(\frac{x}{n}\right) = \frac{1}{n^2} v(x)$$
$$= \frac{1}{n^2} \cdot nPQ$$
$$= \frac{PQ}{n}$$

$$SE(P) = \sqrt{\frac{PQ}{n}}$$

$$z = \frac{t - E(t)}{SE(t)} = \frac{p - E(P)}{SE(P)}$$
$$= \frac{p - P}{\sqrt{PQ/n}}$$

if population proportion P is not known then P approximetly equal to p.

Q.1) A coin was tossed 400 times & head turn up 216 times. Test the hypothesis that coin is unbiased.

$\longrightarrow$ Let the null hypothesis $H_0 : P = 0.5$

$\frac{1}{2} \times \frac{1}{2}$ Alternative hypothesis $H_1 : P \neq 0.5$

$\frac{1}{2} / 2$ (It is two-tailed test).

$= 1/2$ For 5% level of significance $z_\alpha = 1.96$

To find z:

$$z = \frac{t - E(t)}{SE(t)}$$

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

given : $n = 400$
$X = 216$

$$p = \frac{216}{400} = 0.54$$

$$z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = \frac{0.04}{\sqrt{6.25 \times 10^{-4}}}$$
$$= \cancel{6.4 \times 10^{-3} \times 10^{4}}$$
$$= \cancel{64} \ \cancel{0} \ 1.6$$

$z = 1.6 < z_\alpha = 1.96$

$\therefore$ Hypothesis is accepted at 5% level of significance.

$\therefore$ We conclude that coin is unbiased as at 5% ~~of~~ level of significance.

Q.2) ~~In cu~~ If cubical dice is thrown 9000 times & a throw of 3 or 4 is observed ~~310~~ 3440 times. show that the dice cannot be regarded an unbiased one & find extreme limits the probability of throw lies.

$\rightarrow$ H_0 : dice is unbiased. i.e. $H_0 : P = 1/3$

~~H_0 :~~

$\frac{1}{36} + \frac{1}{6} = 1/3$.

$\therefore$ H_1 is $P \neq 1/3$ therefore it is two-tailed test.

To test for 5% level of significance we compare z with $z_\alpha = 1.96$

To find $z = \frac{p - P}{\sqrt{PQ/n}}$

$n = 9000$ $\quad X = 3240$ $\quad P = 1/3$

$$p = \frac{3240}{9000} = 0.36 \qquad Q = 2/3$$

2.469×10^{-5}

0.332

0.22

$$= \frac{0.36 - 1/3}{\sqrt{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)/9000}}$$

$$= \frac{\cancel{0.03}\ 26.66}{\sqrt{2.46 \times 10^{-5}}}$$

$$= \frac{\cancel{0.03}\ 26.66}{4.95 \times 10^{-3}}$$

$$Z = 5.36$$

$\cancel{-1.96 <}$ $Z > 1.96$ Ho is rejected & H_1 is accepted. The dics cannot be regarded as unbiased dics

To find 95% of confidence limit of the proportion.
The confidence limit are given by

$$\left(P - 1.96\sqrt{\frac{PQ}{n}},\ P + 1.96\sqrt{\frac{PQ}{n}}\right)$$

$$\left(\frac{1}{3} - 1.96\sqrt{\frac{1/3 \cdot 2/3}{9000}},\ \frac{1}{3} + 1.96\sqrt{\frac{1/3 \cdot 2/3}{9000}}\right)$$

$$(0.3235,\ 0.9430)$$

Q.3) A manufacture claims that at least 95% equipment which he supplied to a factory confirm to the specification. An examination of a sample of 200 pieces of the equipment rebulid that 18 where faulty. Test this claims at a significant level of 0.05 & 0.01.

$\frac{95}{100}$

H_0 is $P \geq 0.95$

$H_0 : P = 0.95$

$\therefore H_1$ is $P < 0.95$ (left-tailed test)

To test for 5% level of significance

$Z_\alpha = \cancel{1.465}\ -1.695$

To find Z :

$n = 200,\ x = 200 - 18 = 182$

$p = \frac{182}{200} = 0.91$

$P = 95\%$ (success probability)

$= 0.95$

$Q = 0.05$

$$Z = \frac{p - P}{\sqrt{PQ/n}}$$

$$= \frac{0.91 - 0.95}{\sqrt{\frac{0.95(0.05)}{200}}}$$

$$= \frac{-0.04}{1.541 \times 10^{-2}}$$

$$= \frac{4}{1.5}$$

$$= -2.59$$

$Z = -2.597 < Z_\alpha = -1.695$

H_0 is rejected & H_1 is accepted.

We conclude that manufacture's claim is rejected at 5% level of significance.

$z = -2.597 < z_\alpha = -2.33$ for 1% level of significance.

$\therefore$ We conclude that manufactures claim is rejected at 1% of significance.

Test for difference between two sample proportion : — ($t = p_1 - p_2$)

Let two samples of size 's', n_1 & n_2 be drawn from 2 populations with population proportion 'P' with proportions ~~p_1 &~~ P_1 & P_2 resp. Let x_1 & x_2 be the no. of success & p_1, p_2 observed proportions of sample resp.

Then $p_1 = \dfrac{x_1}{n_1}$

$p_2 = \dfrac{x_2}{n_2}$

$Q = 1 - P_1$ & $Q_2 = 1 - P_2$

then we have

$E(p_1) = P_1$

$E(p_2) = P_2$

and

$V(p_1) = \dfrac{P_1 Q_1}{n_1}$

$V(p_2) = \dfrac{P_2 Q_2}{n_2}$

Since, for large samples p_1 & p_2 are normally distributed. Their difference is also normally distributed.

$$E(p_1 - p_2) = E(p_1) - E(p_2) = P_1 - P_2$$

$$V(p_1 - p_2) = V(p_1) + V(p_2)$$

$$= \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}$$

Standard Error:

$$SE(p_1 - p_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

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

$$= \frac{(p_1 - p_2) - E(p_1 - p_2)}{SE(p_1 - p_2)}$$

$$= \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

We set up the null hypothesis that there is no significant difference betn 2 ~~propo~~ population proportion i.e. $H_0: P_1 = P_2 = P$

$$Z = \frac{\cancel{p_1 - p_2}\ p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

If population proportions & not provided then it is obtained using formula

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

If sample proportions is same then

$H_0 : P_1 = P_2$

then $z = \dfrac{P_1 - P_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$

Q. In a sample of 600 men from a certain city 450 is found to be smokers. In another sample of 900 men from another city 450 is smokers. Do the data indicates that cities is significantly different with respect to smoking among men.

→ let H_0 : There is no significant difference with respect to habbit of smoking among the men of 2 cities. ($P_1 = P_2$)

$H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$ (i.e. Right tailed test)

Given n_1, n_2, x_1, x_2

$n_1 = 600$ $x_1 = 450$

$n_2 = 900$ $x_2 = 450$

$P_1 = \dfrac{x_1}{n_1} = 0.75$

$P_2 = \dfrac{x_2}{n_2} = 0.5$

$P = \dfrac{P_1 n_1 + P_2 n_2}{P_1 + P_2} = \dfrac{0.75 \times 600 + 0.5 \times 900}{0.75 + 0.50}$

$P = 0.6$

0.6648×10^{-3} 0.25 2.57×10^{-2}

$\sqrt{0.24(\quad)}$

$1.66 \times 10^{-3} +$
1.11×10^{-3}
2.77×10^{-3}

$Q = 1 - P = 1 - 0.6 = 0.4$

$$z = \frac{\cancel{P_1 - P_2}\; P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

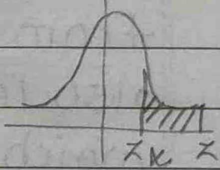
$$= \frac{0.75 - 0.50}{\sqrt{(0.6)(0.4)\left(\frac{1}{600} + \frac{1}{900}\right)}}$$

$$= \cancel{9.727}\; 9.682$$

For 5 % level of significance the critical value for one tailed test is $z_\alpha = 1.645$

$z_\alpha = 1.645 < z = 9.682$

H_0 is rejected.



For 1 % level of significance, the critical value is $z_\alpha = 2.33$

$z_\alpha = 2.33 < z = 9.682$

H_0 is rejected.

$\therefore$ There is significant difference with respect to habbit of smoking among the ~~ma~~ men of two cities.

Test for difference between two sample proportion :

Let two samples of sizes n_1 & n_2 be drawn from two populations with population proportion P_1 & P_2 resp. also let x_1 & x_2 be the number of successes & p_1 & p_2 be observed proportion of sample resp. then,

Test for difference between sample mean & population mean :

Let a sample of size 'n' be draw from population with mean μ & varience δ^2 also let sample have observation $x_1, x_2, ---, x_n$ which & independent & identically distributed then each x_i is an independent normal variate with mean μ_i & varience δ_i^2.

We have the sample mean $\overline{x} = \frac{1}{n} \Sigma x_i$

& the sample varience $s^2 = \frac{1}{n} \Sigma (x_i - \overline{x})^2$

$$E(\overline{X}) = E\left(\frac{1}{n} \Sigma x_i\right)$$

$$= \frac{1}{n} E(\Sigma x_i)$$

$$= \frac{1}{n} \Sigma \mu_i$$

$$= \frac{1}{n} n \cdot \mu$$

$$E(\overline{X}) = \mu$$

$$v(\overline{x}) = v\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{1}{n^2}\left[v(x_1) + v(x_2) + \cdots + v(x_n)\right]$$

$$= \frac{1}{n^2} n\delta^2$$

$$v(\overline{x}) = \frac{\delta^2}{n}$$

$$SE(\overline{x}) = \sqrt{\frac{\delta^2}{n}} = \frac{\delta}{\sqrt{n}}$$

$$\left| SE(\overline{x}) = \frac{\delta}{\sqrt{n}} \right|$$

$$z = \frac{t - E(t)}{SE(t)} = \frac{\overline{x} - E(\overline{x})}{SE(\overline{x})}$$

$t =$ mean $\left| z = \frac{\overline{x} - u}{\delta/\sqrt{n}} \right|$

If δ i.e. population deviation is that we use $\delta = s$ i.e. sample stand-deviation.

The confidence limit for population mean are given

$$t \pm z\alpha \, SE(t)$$

$$P \pm z\alpha \sqrt{\frac{PQ}{n}}$$

$$\overline{x} \pm z\alpha \frac{\delta}{\sqrt{n}}$$

1) The mean weight obtain form random sample of size 100 is 64 the standerd deviation of the weight distribution of population is 3g. Test the statement that mean weight of population is 67g for 5% level of significance Also, setup 99% confidence limit of mean weight of population.

→ Given :

$n = 100$

$\bar{x} = 64g$

$\delta = 3g$ $\qquad$ $\mu = 67g$

$H_0 : \mu = 67g$

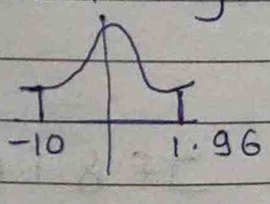
$H_1 : \mu \neq 67g$ (two-tailed test)

$$Z = \frac{\bar{x} - \mu}{\delta / \sqrt{n}} = \frac{64 - 67}{3/\sqrt{100}} = \frac{-3}{0.3}$$

$Z = -10$

Critical value of z at 5% level of significance for two tailed is 1.96

$Z = -10 < Z_\alpha = 1.96$



H_0 is rejected at 5%. Level of significance & we conclude that sample is not drawn from with population with mean 67.

To find confidence limit 99%.

The critical value for 1% level of significance is 2.58

∴ The confidence limit are

$$\bar{x} \pm Z_\alpha \frac{\delta}{\sqrt{n}}$$

$$= 64 \pm 2.58 \left(\frac{3}{\sqrt{100}} \right)$$

$= 64 \pm 0.774$

$= (64 + 0.774)$ or $(64 - 0.774)$

$= 64.774$ or 63.226

$= (64.774, 63.226)$