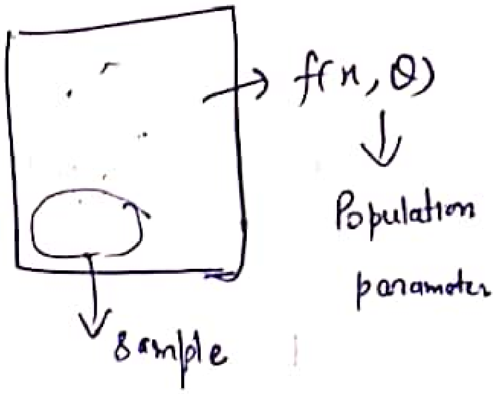
Testing of Hypothesis.



Def^n :- The assumption about population parameter is known as statistical hypothesis.

Hypothesis :- Quantitave statement about population

e.g. Average weight of SVNIT student is 50

i.e. $\mu = 50$ kg.

Question ?

Is above statement valid ? For this we have to test the hypothesis.

Null Hypothesis : We are testing the null hypothesis for possible ~~assumption~~ rejection under assumption that it is initially true. (H_0)

Alternative Hypothesis : Complementary of null hypothesis is known as alternative hypothesis. (H_1)

How we will set up the hypothesis

(Ex) A manufacturer company of cold drink claiming that they are putting 500 ml in each cold drinks bottel

$H_0 : \mu = 500 ml$

$H_1 : \mu \neq 500 ml$ [Two tailed test]

or $H_1 : \mu < 500 ml$ [One-tailed test (left tailed test)]

or $H_1 : \mu > 500 ml$ [One tailed test] (Right tailed test)

Note: We will always performs the test against null hypothesis. If evidence is against it we will reject the null hypothesis. If evidence is in favor of this then we will not reject null hypothesis.

Note: We will never write that we will accept null hypothesis because we are performing the test based on sample only.

Statistical test: Statistical test are conducted to test the hypothesis and to find the inference about population parameter.

(i) Parametric tests (ii) Non-parametric test.

(i) Defⁿ: Parametric tests are applied under the circumstances where population is normally distributed or is assumed to be normally distributed.

e.g T-test, Z-test, F-test.

Note: These tests are applied where the data is quantitative.

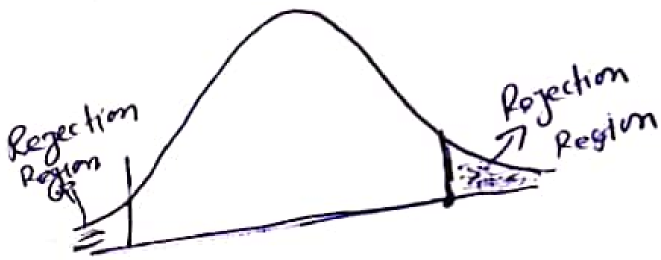
(ii) Defⁿ: Non-parametric tests are applied under the circumstances where the population is not normally distributed (or skewed distributed).

* These tests are also called distribution free tests e.g Chi-square test.

T-test: This is based on t-distribution.

It is used to test the significant difference of mean values when sample size is small (<30) and population standard deviation is not available.

$$t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$$



$\bar{x} \rightarrow$ sample mean

$\mu \rightarrow$ population mean

$n \rightarrow$ sample size

$s \rightarrow$ sample standard deviation.

Decision: If calculated value of t > tabulated value of t, we reject the null hypothesis.

If $t_{cal} < t_{tab}$, then we do not reject the null hypothesis.

Z- Test :

It is used to determine whether the means are significant when the population variance is known and the sample size is large ($\because n > 30$).

$$z = \sqrt{n}\left(\frac{\bar{x} - \mu}{\sigma}\right)$$

$\bar{x} \to$ sample mean

$\mu \to$ population mean

$\sigma \to$ standard deviation of population

$n \to$ sample size ($n > 30$).

Note: ① Sample size is large and the population variance is not known $\to$ Z- test.

② Sample size is small and the population variance is known $\to$ Z- test.

F - Test: It is test for the null hypothesis that two normal populations have the same variance.

An F-test is regarded as a comparison of equality of sample variances.

F-test is simply a ratio of two variances.

$$F = \frac{S_1^2}{S_2^2}$$

Chi - Square Test: It is non-parametric test.

Chi-square test can be used (i) as a test of goodness of fit

(ii) To test of independence between two variables.

Parametric test

$$\chi^2 = \frac{\Sigma(O-E)^2}{E} \quad \boxed{= \frac{(n-1)S^2}{\sigma^2}}$$

Scanned with CamScanner

⑦

Level of Significance : Probability of rejecting null hypothesis when it is true (error).

Q: A random sample of 50 items gives the mean 6.2 and variance 10.24. Can it be regarded as drawn from a normal population with mean 5.4 at 5% level of significance

Sol:

$\bar{x} = 6.2$

$s = \sqrt{10.24} = 3.2$

$\mu = 5.4$

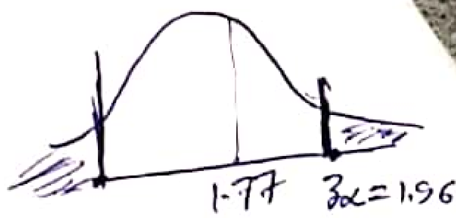
Null hypothesis $H_0 : \mu = 5.4$

Alternate hypothesis $H_1 : \mu \neq 5.4$

Use Z-test

$$Z_d = \frac{6.2 - 3.2}{3.2/\sqrt{50}} = 1.77$$

$Z_{tabulated} = 1.96$



Since $Z_{calculated} < Z_{tabulated}$

then we will not reject the null hypothesis.

Ex: A random sample of 400 members is found to have a mean of 4.45 cm. Can it be reasonably regarded as a sample from a large population, whose mean is 5 cms and variance is 4 cm.

Sol^n

$n = 400, \quad \bar{x} = 4.45$

$\mu = 5, \quad \sigma^2 = 4$

Null hypothesis $H_0 : \mu = 5$

Alternative hypothesis $H_1 : \mu \neq 5$

$; \alpha = 5\%$

$$Z_{cal} = \left| \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right| = 5.5, \quad Z_\alpha = 1.96$$

$Z_{cal} > Z_\alpha$, We reject the null hypothesis.

Note: When we are testing Null hypothesis (H_0) against alternative hypothesis (H_1) there are four possibilities

H_0 accepted when H_0 is true

H_0 rejected when H_0 is true $\rightarrow$ Type I error

H_0 accepted when H_0 is false $\rightarrow$ Type II error

H_0 rejected when H_0 is false

Type I error

$\alpha = P(\text{Type I error})$

$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$

$\alpha \rightarrow$ Size of ~~test~~ type I error

$\alpha = P\{x \in W \mid H_0\}$

Type II error

$\beta = P(\text{Type II error})$

$\beta = P(\text{Accept } H_0 \mid H_1 \text{ is true})$

$1-\beta \rightarrow$ power of test.

$\beta \rightarrow$ size of type-II error

$\beta = P(x \in W^c \mid H_1)$

(Ex) Given the frequency distⁿ,

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} , & 0 \le x \le \theta \\ 0 ; & \text{otherwise} \end{cases}$$

and that you are testing the null hypothesis $H_0: \theta = 1$ against $H_1: \theta = 2$ by means of a single observed value of x. What would be the size of the type-I and type II errors, if you choose the interval (i) $0.5 \le x$ ~~(ii)~~ as the critical regions? Also obtain the power function of the test.

Solⁿ: Here we want test $H_0: \theta = 1$ against $H_1: \theta = 2$

(i) $W = \{x : 0.5 \le x\} = \{x : x \geq 0.5\}$

$\bar{W} = \{x : x < 0.5\}$

$\alpha = P(x \in W \mid H_0)$

$= P(x \geq 0.5 \mid \theta = 1) = P(0.5 \le x \le \theta \mid \theta = 1)$

$= \int_{0.5}^{\theta} \frac{1}{\theta} dx = \frac{1}{\theta}(\theta - 0.5)$

$= 0.5$

$$\beta = P(x \in W^c \mid H_1)$$

$$= P(x \le 0.5 \mid \theta = 2)$$

$$= P(0 \le x \le 0.5 \mid \theta = 2)$$

$$= \int_0^{0.5} \frac{1}{\theta}\, dx = \frac{1}{2}(0.5) = 0.25$$

Thus the size of type I ~~error~~ and type II errors are respectively $\alpha = 0.5$ and $\beta = 0.25$

and power of the test $= 1 - 0.25 = 0.75$.