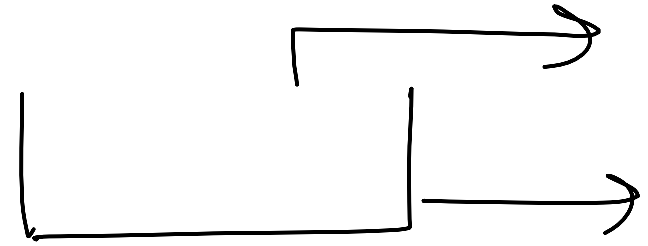
Example:-



Urn

It has 10 balls.
(containing white & red balls).

Let p be the prob. that is related to having white balls in a particular draw.

The issue of concern:- By drawing a ball 10 times, we want to decide no. of ~~drawn~~ white balls 5 or 7.

Hypothesis:- $H_0: p = 0.5$ against $H_1: p = 0.7$

Null hypothesis

Alternative hypothesis.

Supervised learning technique → Rule : Accept H_0 if $X < 6$ — $\circledast$

X → Total no. of white balls.

If $X = 2$, then H_0 will be accepted according to $\circledast$

If $X = 8$, " H_0 " " rejected " " $\circledast$.

Question :- The general rule :- "Accept H_0 if $X < c$".

i) How to choose c??

ii) What will be the form of acceptance or rejection region :- (why not $|X| > d$)

Basic terminology :-

i) Hypothesis :- Null & Alternative.

ii) Test Statistic :- The test will be carried out ~~the~~ based on the test statistic.

iii) Rejection region :- The ~~reject~~ region, where H_0 will be rejected.

iv) Value of the test statistic for a given data (Required to compute the p-value).

v) Final decision :- Reject H_0 or accept H_0.

Rejection region :- It is a subset of the sample space that corresponds to rejection of H_0. Sometimes, it is called critical region.

A few fundamental quantities :- Type-I error, Type-II error & Power of the test.

True situation

Decision	H_0 is true	H_0 is NOT true.
H_0 is not rejected	Correct	Type-II error,
H_0 is rejected	Type-I error	Correct

The person is innocent

The person is criminal

The person is proven innocent

The person is proven guilty

P[Type-I error] should be very small & fixed.

Power of the test

$= P[H_0 \text{ is rejected} \mid H_0 \text{ is NOT true}]$

Test-Statistic:- A statistic $T_n = T_n(X_1, \cdots, X_n)$ is called a test-statistic if the dist^n of T_n under H_0 is independent of the unknown parameter θ.

Example:- X_1, X_2 i.i.d. with p.m.f. $f(x;\theta)$, $\theta = 0$ or 1.

$H_0: \theta = 0$ against $H_1: \theta = 1$.

	$x=0$	$x=1$
$f(x;\theta=0)$	0.2	0.8
$f(x;\theta=1)$	0.6	0.4

Rule:- H_0 will be rejected if $X_1 + X_2 < 2$.

$$P[\text{Type I error}] = P[\text{Reject } H_0 \mid H_0 \text{ is true}].$$

$$= P[X_1 + X_2 < 2 \mid \theta = 0].$$

When $X_1 + X_2 < 2$?? if & only if

$(X_1 = 0 \ \& \ X_2 = 0)$ & $(X_1 = 1 \ \& \ X_2 = 0)$ & $(X_1 = 0 \ \& \ X_2 = 1)$

$$P[X_1 + X_2 < 2 \mid \theta = 0] = P[X_1 = 0, X_2 = 0 \mid \theta = 0]$$
$$+ P[X_1 = 1, X_2 = 0 \mid \theta = 0]$$
$$+ P[X_1 = 0, X_2 = 1 \mid \theta = 0]$$

$X_1 \perp\!\!\!\perp X_2$

$$= P[X_1 = 0 \mid \theta = 0]\, P[X_2 = 0 \mid \theta = 0] + P[X_1 = 1 \mid \theta = 0]\, P[X_2 = 0 \mid \theta = 0]$$
$$+ P[X_1 = 0 \mid \theta = 0]\, P[X_2 = 1 \mid \theta = 0]$$

= Computing it = (follows from table).

P[Type II error] = P[Accept Ho | Ho is NOT true]

$$= P[X_1 + X_2 \geq 2 \mid H_0 \text{ is NOT true}].$$

$$= P[X_1 = 1,\ X_2 = 1 \mid \theta = 1].$$

$$\overset{X_1 \perp\!\!\!\perp X_2}{=} P[X_1 = 1 \mid \theta = 1]\ P[X_2 = 1 \mid \theta = 1]$$

= Compute it (Take help from table).

Power of the test :- 1 − P[Type-II error]

= compute]

P[Ho is rejected | Ho is NOT true]

check it.

$$= P[X_1 + X_2 < 2 \mid \theta = 1] = P[X_1 = 0, X_2 = 0 \mid \theta = 1] + P[X_1 = 1, X_2 = 0 \mid \theta = 1] + P[X_1 = 0, X_2 = 1 \mid \theta = 1]$$

(last year, got the answer = 0.64).

Discussion:- **Simple hypothesis:-** A hypotheses that completely specify the $dist^n$ for a given data.

Composite hypothesis:- A hypothesis, which is not simple (i.e, it gives you the Class of $dist^n$).

Recall the Urn example:-

Decision was that Reject H_0 if $X > c$.

Suppose $\alpha = 0.06$ is given to you.

$\hookrightarrow P[\text{Type-I error}]$

$$P[\text{Type - I error}] = P[H_0 \text{ is rejected} \mid H_0 \text{ is true}].$$

$$= P[H_0 \text{ is rejected} \mid p = 0.5].$$

Using binomial prob.

$$P[X \geq 7 \mid p = 0.5] = 0.1719$$

$$P[X \geq 8 \mid p = 0.5] = 0.0547.$$

$$P[X \geq 9 \mid p = 0.5] < 0.0547.$$

<u>Decision:-</u> Reject H_0 if $X \geq 8$ at $\alpha = 0.06$.

Sometimes it is called the level of significance of the test.