Non-Parametric tests :

(Parametric Tests) – Covered so far

Test-statistic

$$- \frac{\bar{x} - c}{\sigma/\sqrt{n}} \; ; \; \frac{\bar{x} - c}{s/\sqrt{n}} \; ; \dots$$

– to perform the test assumption of Normal distribution or by assuming C.L.T. Holds

$\boxed{\text{Sample size being large}}$

Non-parametric tests :-

$\boxed{\text{No need to Device a test for normality}}$

- Distribution free tests
- less powerful than parametric tests

  Power of test :- next statistics course
- Very helpful for a start.

Example :- – Vaccine has been developed

Q:- Is the vaccine effective or not? (for a disease)

A:- Designed an experiment

– choose n individuals from a population

- gave vaccine to n_1 of them
placebo to $n_2 := n - n_1$

	Affected	Not affected	Total
vaccine	X_{11}	X_{12}	n_1
placebo	X_{21}	X_{22}	n_2

$\tilde{Q}$: Based on table can we answer the question Q ?

Example 2 :- [Tea-Tasting Example]

A - claimed :- Can tell from tasting a cup of Tea whether :-

$\boxed{\text{English preparation}}$
- milk first and Tea next
- Tea first and milk next

Deviced an Experiment :-

- Prepared 8 cups of Tea
  - 4 milk first
  - $\boxed{4}$ Tea first

- Person A tasted each one of them and gave opinion.

	Tea	Milk	Total
Tea first	3	1	4
Milk first	1	3	4

Approach 1 (Parametric) : χ^2- test for independence.

Approach 2 :-

Experiment:

	Tea	Milk	Total
Tea first	$3 \equiv X_{11}$	1	4
Milk first	1	3	4

- Row totals are fixed by the experimenter
- Column totals are decided by person A

H_0 :- Person A has No ability
(i.e. Person A calls "tea first" by choosing a random Sample from 8)

Under H_0 :-

$\mathbb{P}(X_{11} = 3) = \mathbb{P}($ choose 4 cups from 8 cups & 3 of them are actually Tea 1st $)$

(Hypergeometric) $= \dfrac{\binom{4}{3}\binom{4}{1}}{\binom{8}{4}} = 0.229$

Test p-value :- $\mathbb{P}(X_{11} \geq t_0)$ observed

Specify level of significance := $\alpha = 0.05$

Here $t_0 = 3$:- Assume H_0 is true

$$\mathbb{P}(X_{11} \geq 3) = \mathbb{P}(X_{11} = 3) + \mathbb{P}(X_{11} = 4)$$

Under H_0 :- $X_{11} \sim$ Hypergeometric $(N = 8, G = 4, n = 4)$

$$\mathbb{P}(X_{11} \geq 3) = 0.229 + \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}}$$

$$= 0.229 + 0.014$$

$$= 0.243$$

$\therefore$ As $\mathbb{P}(X_{11} \geq 3) = 0.243 \gg \alpha = 0.05$

$\therefore$ the null cannot be rejected.

- there is not enough evidence to reject the null hypothesis that person A was purely guessing

---

<u>Sign test</u> & <u>Signed Rank tests</u>

<u>Model</u> : $X_1, \ldots, X_n$ are from a random sample

$$X_i = \theta + e_i$$

Errors independent with p.d.f $f(\cdot)$ symmetric around 0.

$$H_0: \theta = 0 \qquad H_A: \theta > 0$$

Test statistic

$$S = \sum_{i=1}^{n} \operatorname{sgn}(X_i)$$

where
$$\operatorname{sgn}(t) = \begin{cases} -1 & t < 0 \\ 0 & t = 0 \\ 1 & t > 0 \end{cases}$$

Sign test

- look at positive observations

$$S^+ = \#\{i : X_i > 0\}$$

- ignore 0's and sample is reduced.

- If we observe -1 and 1 then

X_i - can be thought as Bernoulli(·)

$\Rightarrow$ $S^+ \sim$ Binomial distribution $\left(n, \frac{1}{2}\right)$ $\equiv$ Null distribution

<u>Test</u> :- Compare this distribution with the observed statistic

Example (Sign test) :-

12 people are chosen

10 prefers $\underbrace{\text{shorts over full pants}}_{(X_i > 0)}$

1 prefers full pants over shorts $(X_i < 0)$

1 no preference $(X_i = 0)$

- strong preference for shorts ?

How likely is such a result true if

H_0: there is no preference over shorts or full pants

is true ?

$S^+ = \{i : X_i > 0\} \sim \text{Bin}(n=11, p=\frac{1}{2})$

$s^+ \equiv$ observed $= 10$

p-value :- $\mathbb{P}(S^+ \geq 10) \cong 0.0059$

α - level of significance :- 0.05

$\Rightarrow$ Since $0.0059 \ll 0.05$

H_0: No preference H_A: shorts over full

we reject null hypothesis.

Signed Rank - Wilcoxon Test

(Compare with t-test)

$H_0: \mu = 0$, $H_A: \mu > 0$

Not distribution free

Test-statistic : $t_0 := \frac{\overline{X}}{S/\sqrt{n}} \sim t_{n-1}$

p-value : $\mathbb{P}(t_{n-1} > t_0)$

$X_1, X_2, \dots\dots, X_n$ - sample $\left\{ \begin{array}{l} X_i = \theta + e_i \\ \text{(old model)} \end{array} \right.$

$$W = \sum_{i=1}^{n} \operatorname{Sgn}(X_i) \operatorname{Rank}(|X_i|)$$

<u>Test- statistic</u>

$$W^+ = \sum_{i=1}^{n} \operatorname{Rank}(|X_i|) \mathbb{1}(X_i > 0) \overset{Ex}{=} \frac{1}{2}\left(W + \frac{n(n+1)}{2}\right)$$

W^+ symmetric around $\frac{n(n+1)}{4}$

- There is no closed form for distribution of W^+

$H_0: \theta = 0$ versus $H_A: \theta > 0$

p-value: $\mathbb{P}(W^+ > t_0)$

t_0 - observed test statistic