Sign Test

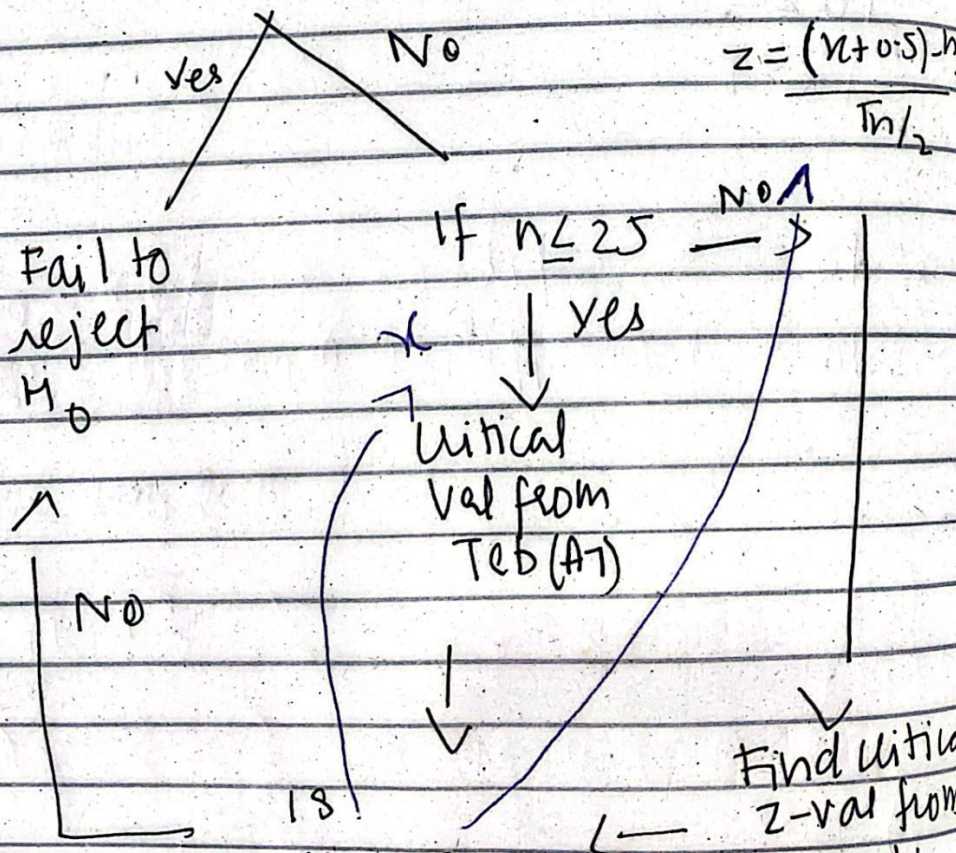
- claims involve matched pair of sample data
- claims involving nominal data
- claims about media of single popn.

Assign +ve & -ve sign

$n \rightarrow$ Total no. of signs

$x \rightarrow$ Less freq sign

if sample contradict with H_1



Claim About Metched pair

Q#1

Before	After	Sign
21	19	+
34	22	+
9	16	−
45	31	+
30	21	+
54	30	+
37	22	+
36	18	+
33	17	+
40	21	+

2

① H_0: There is no Diffu.
H_1: There is diffu

② Signifiqance Lvl $= \alpha = 0.0255$

③ $n = 10$ $x = 1$

④ As $x \leq n$ $(1 \leq 10)$ is True
we will reject H_0

Q # 2 :

Before vaccine	After vacan	Sign
3	2	+
4	1	+
2	0	+
1	1	0
3	1	+
6	3	+
4	3	+
5	2	+
2	2	0
0	2	−
2	3	−
5	4	+
3	3	0
3	2	+

① H_0 : There is no diffu
H_1 : There is a diffu

② $\alpha = 0.05$

③ $n = 12$ $K = 2$ → From table A-7

④ As $2 \leq 2$ so we will reject H_0

Population median

Left tail	Right tail	Two tail
$H_0: \mu \geq K$	$H_0: \mu \leq K$	$H_0: \mu = K$
$H_a: \mu < K$	$H_a: \mu > K$	$H_a: \mu \neq K$

Tip: Assign + if val is > median.
- if < median 0 if = median

−	−	+	+	0
1469	1462	1634	1602	1500
−1463	−1476	+1570	+1544	−1452
1481	+1523	+1525	+1548	1511
+1579	+1620	+1568	−1492	+1649

① $H_0: \mu \leq 1500$

$H_a: \mu > 1500$

② $\alpha = 0.05$

③ $n = 19$ $x = 7$

④ As $7 \leq 4$ which is false
So we will accept H_0

Q#2:

-118 +167 -72 -79 -76 -106 -102 +11

-73 -119 +162 -114 @ 120 -93 +135 +14

-77 +157 -115 -88 +152 -70 -85 -91

① $H_0: \mu \leq 120$

$H_a: \mu > 120$

② $\alpha = 0.025$

③ $n = \cancel{19}23$ $x = 6$

④ As $6 \leq$ ~~16~~ ~~(false)~~ True

so we will ~~accept~~ reject H_0

claims
~~About~~
nominal data

~~H_0~~

① $H_0 : p = 0.5$
$H_a : p \neq 0.5$] ← Two tail test

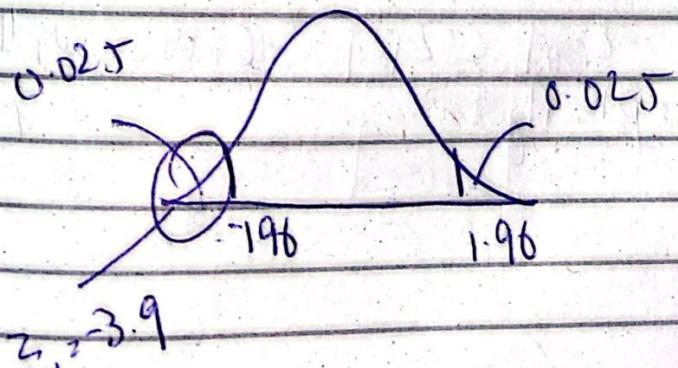
② $\alpha = 0.05$ (~~Tw~~

③ $n = 100$ $x = 30$ (new)

④ As $n > 25$ so will use

$$z = \frac{(x + 0.5) - n/2}{\sqrt{n}/2}$$

$z = -3.9$



So will reject H_0

Treat Right tail as left tail

(2)

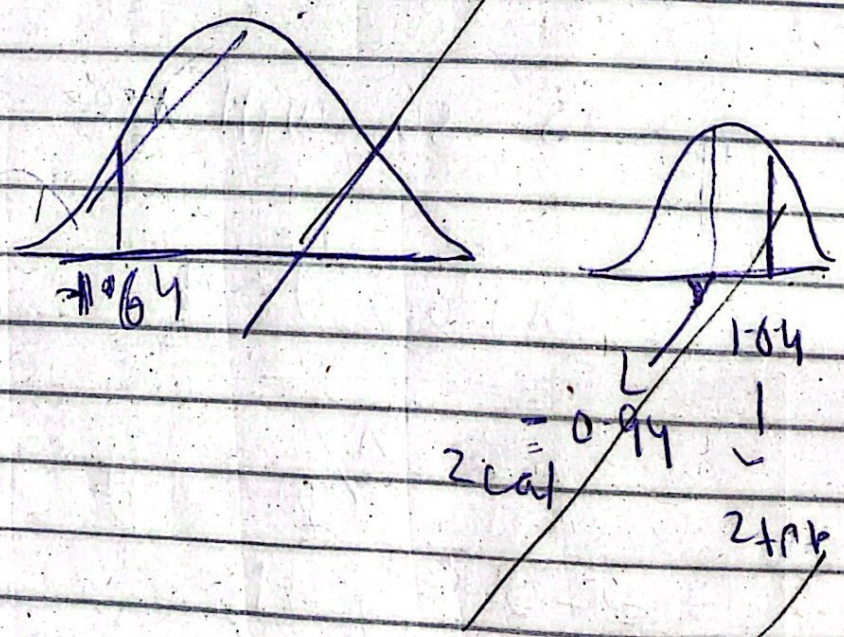
(1) $H_0: p = 0.5$ (was not smoking after year)

$H_a: p < 0.5$

(2) $\alpha = 0.05$

(3) $n = 71$ $x = 31$

$$Z = \frac{(x+0.5) - n/2}{\sqrt{n}/2} = -0.94$$



So,

As $Z_{cal} < Z_{tab}$ so we will accept H_0

③

① $H_0: P <= 0.5$

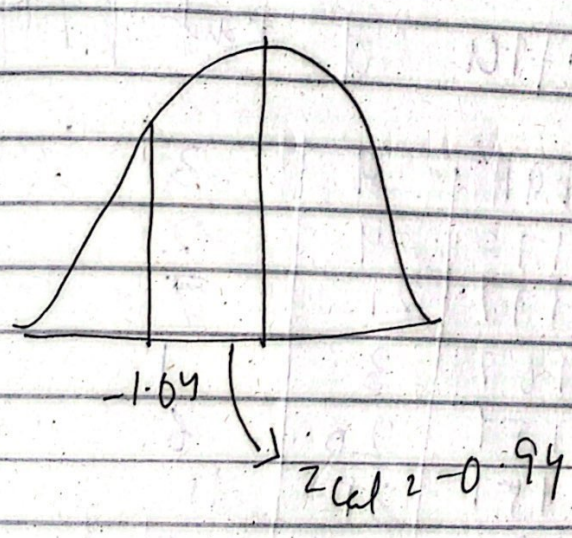
$H_a: P > 0.5$

② $\alpha = 0.05$

② $\alpha = 0.05$

③ $n = 71$ $n = 31$

$z = -0.94$



As, $z_{cal} > z_{tab}$ so we will accept H_0 (Majority of people not smoke after year)

Wilcoxon Signed Rank

If $n \le 30$

x ← test-static

If $n > 30$

$$Z = T - \frac{n(n+1)}{4} \Big/ \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Claim About Matched pair (dependent sample)

Claim about median of single population

Matched pair

Is Drug has effect?

Before	After	d	Rank of |d|	Signed Rank
200	191	9	3	
174	170	4	2	
198	177	21	7	
170	167	3	1	
179	159	20	6	
182	151	31	11	
193	176	17	5	
209	183	26	9.5	
185	159	26	9.5	
155	145	10	4	
169	146	23	8	
210	177	33	12	

Sum of positive ranks = 78

" Negative = 0

T = Smaller (78, 0)

= 0

n = 12

As $n \leq 30$ so

critical_val = 14

$0 \leq 14$ (Reject H_0)

=> Mean There is a difference

Q#2

Sitting	Supine	d	Rank of |d|	Signed Rank
2.96	1.97	0.99	5	
4.65	3.05	1.6	9	
3.27	2.29	0.98	4	
2.50	1.68	0.82	3	
2.59	1.58	1.01	6	
5.97	4.43	1.54	8	
1.74	1.53	0.21	1	
3.51	2.81	0.7	2	
4.37	2.70	1.67	10	
4.02	2.70	1.32	7	

Sum of positive rank = 58

" = 0

T = 0

$0 < 8$ (Reject H_0)

Mean There is a diff

Median of Single population

Day 2 98 97.4 97.8 97 97

Media → 98.6°F

$H_0: \mu = 98.6°F$

$H_1: \mu \neq 98.6°F$

	Media	d	Rank of |d|	Sign Rank
98	98.6	-0.6	1	-1
97.4	98.6	-1.2	3	-3
97.8	98.6	-0.8	2	-2
97	98.6	-1.6	4.5	-4.5
97	98.6	-1.6	4.5	-4.5

Sum of +ve rank = 0

Sum of absolute -ve = 15

$T = 0$

Let $T = 0$ put in formula if $n > 30$

Wilcoxon Rank Sum Test (Independent Samples)

Steps

↓

Determine critical value (z-table) using α

↓

$n_1 \leq n_2$

↓

Find sum of ranks for smaller sample

R

① list combined data in ascending order

② Rank it

Add ranks of smaller sample n_1

↓

$$z = \frac{R - \mu_R}{\sigma_R}$$

↓

Reject or Accept Null Hypothesis

EXP:

$\alpha = 0.10$

Is there a difference b/w earning of male & female?

Male	78	93	114	101	98	94	86	95	117	99	
Female	86	77	101	93	85	98	91	87	84	97	
									100	90	

(1)

H_0: There is no diff

H_a: There is a diff

(2) $\alpha = 0.10$

$n_1 = 10$ $n_2 = 12$

77 (F)	1	94 (M)	12
78 (M)	2	95 (M)	13
84 (F)	3	97 (F)	14
85 (F)	4	98 (M)	15.5
86 (M)	5.5	98 (F)	15.5
86 (F)	5.5	99 (M)	17
87 (F)	7	100 (F)	18
90 (F)	8	101 (M)	19.5
91 (F)	9	101 (F)	19.5
93 (M)	10.5	114 M	21
9 (F)	10.5	117 M	22

(4) R = Sum of rank of sample n_1

$$R = 2 + 5.5 + 10.5 + 12 + 13 + 15.5 + 17 + 19.5 + 21 + 22$$

$$= 138$$

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

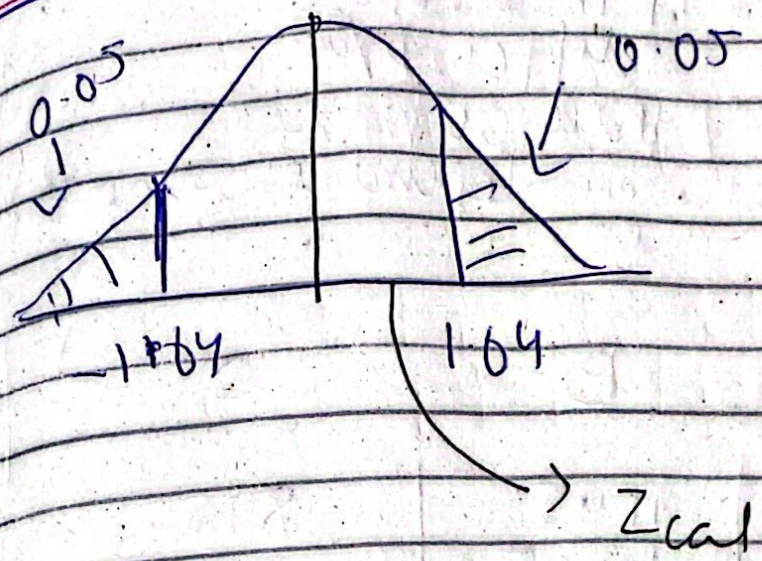
$$= \frac{10(23)}{2} \qquad \mu_R = 115$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$\approx 15.17$$

$$\cancel{z = 138}$$

$$z_{cal} = \frac{R - \mu_R}{\sigma_R} = \frac{138 - 115}{15.17} \approx 1.51$$



~~As $Z_{cal} \not< Z_{tab}$~~ so there

As Z_{cal} lies in rejection region
we fail to reject H_0

Kruskal-Wallis

→ If there a diffr b/w three or more independent sample

A	B	C
930	1105	818
725	1025	730
890	980	546
1040	1295	625
1165	1110	912
806	889	618
947	751	535

① H_0: There is no diffy
H_a: There is "

② $\alpha = 0.05$

③ $df = 2$

④ $H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} - \frac{R_k^2}{n_k} \right)$

Ordered Data		Rank
535	(3)	1
546	(3)	2
618	(3)	3
625	(3)	4
725	(1)	5
730	(3)	6
757	(2)	7
806	(1)	8
885	(3)	9
889	(2)	10
890	(1)	11
912	(3)	12
930	(1)	13
947	(1)	14
980	(2)	15
1025	(2)	16
1040	(1)	17
1105	(2)	18
1110	(2)	19
1165	(1)	20
1295	(2)	21

$N = 21 \quad n_1 = 7 \quad n_2 = 7 \quad n_3 = 7$

$R_1 = 5 + 8 + 11 + 13 + 14 + 17 + 20$

$R_1 = 88$

$R_2 = 7 + 10 + 15 + 16 + 18 + 19 + 21$

$= 85 + 21 = 106$

$R_3 = 9+6+12+9+12+5+1$

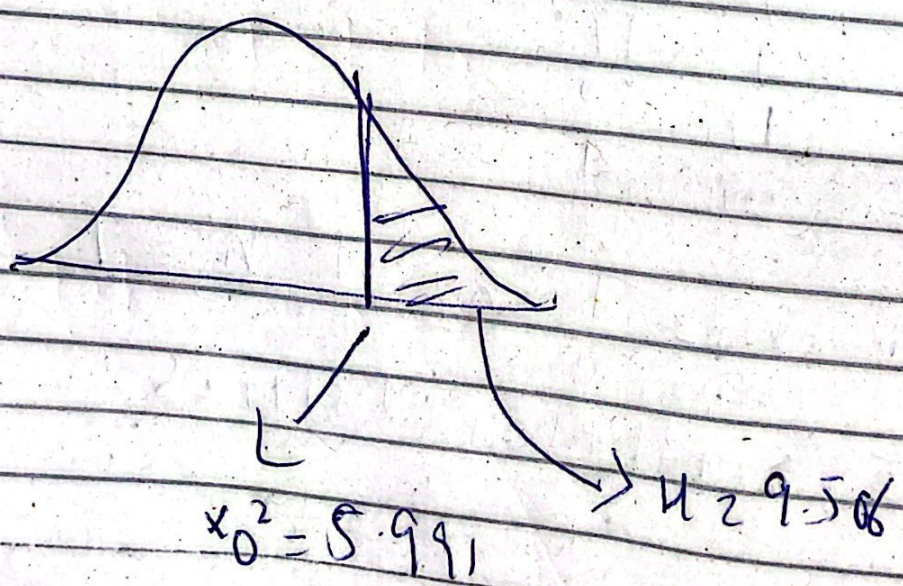
$= 37$

$$H = \frac{12}{21(21+1)}\left(\frac{88^2}{7} + \frac{106^2}{7} + \frac{37^2}{7}\right) - 3(21+1)$$

$H = 75.50 - 3(21+1)$

$H = 9.506$

Critical value $= 5.991$



So, we will reject H_0

Rank correlation

→ Non parametric Test to check association b/w two variables

Are data in form of rank

↓

If not convert it

2(1) 3(4) 5(3)
6(2) 8(3) 2(1)

↓

If there is tie among rank of either variable

2(1.5) 3(3) 5(1.5)

yes → cal d

NO → cal d (ignore sign)

$$r_s = \frac{n\sum xy - (\sum x)\sum y}{\sqrt{n(\sum x)^2 - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Is $n < 30$

Yes

critical value of from table

$t_s = \frac{\bar{x}}{\sqrt{n-1}}$

No

calculate cor

n x

Runs Test → If there is random nos in data.

n_1 → No. of elements of first type

n_2 → No. of elements of second type

r → No. of runs

Is $n_1 > 20$ — yes → $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$

yes ↓

Is $n_2 > 20$ — yes → $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$

NO ↓

$\sigma_r^2 = \ldots$

↓

$z = ?$

↓

Get critical val of z from table

NO ↓

Is $\alpha = 0.05$

yes ↓

Test stat is r

↓

Critical value from table