

# Wilcoxon Rank-Sum Test

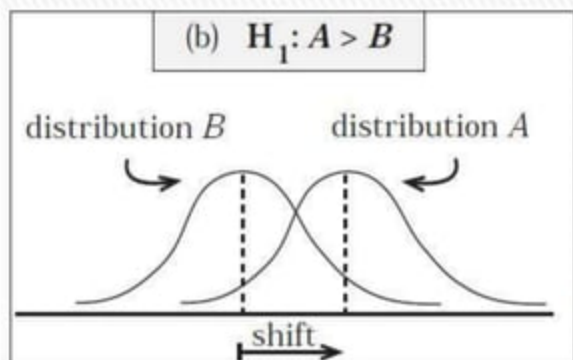
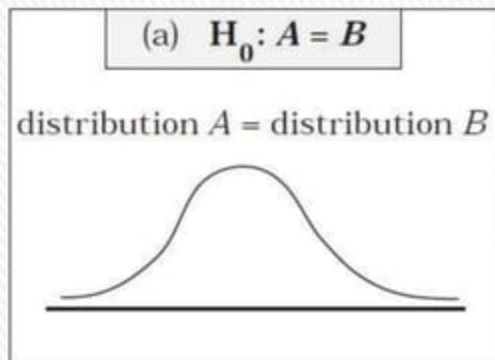
Presentation by:

*Sahil Jain*

*IIT-Delhi*



- » Generally used when normality assumption for the sample does not hold and sample size is small
- » Non-parametric statistical hypothesis test for assessing whether one of two samples of independent observations tends to have larger values than the other.



# Steps

- » Select a random sample from each of the populations
- » Let  $n_1$  and  $n_2$  be the number of observations in the smaller and larger sample respectively
- » Arrange the combined  $n_1+n_2$  observations in ascending order and substitute a rank of 1, 2... to the  $n_1 + n_2$  observations
- » In the case of ties, we assign the conflicting observations with their mean ranks



»  $w_1$ : sum of the ranks corresponding to  $n_1$  observations

»  $w_2$ : sum of the ranks corresponding to  $n_2$  observations

$$» u_1 = w_1 - \frac{n_1(n_1+1)}{2}$$

$$» u_2 = w_2 - \frac{n_2(n_2+1)}{2}$$



- » Our decision is based on the value of the test statistic  $U$  (the random variable for  $u$ )
- » For one-tail test:  $u_1$  or  $u_2$
- » For two-tail test:  $u = \min(u_1, u_2)$

$H_0$	$H_1$	Compute
	$\mu_1 < \mu_2$	$u_1$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$u_2$
	$\mu_1 \neq \mu_2$	$u$

- » Null hypothesis will be rejected whenever the appropriate statistic  $U_1$ ,  $U_2$  or  $U$  assumes a value less than or equal to the desired critical value.

Test Statistic  $\leq$  Critical Value



# Example (two tail)

Brand A	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3		
Brand B	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	5.4

The nicotine content of two brands of cigarette, measured in mg, was found to be as given the table. Test the hypothesis, at 0.05 level of significance, that the median nicotine content of two brands are equal against the alternative that they are unequal.



»  $H_0: \mu_1 = \mu_2$

»  $H_1: \mu_1 \neq \mu_2$

»  $\alpha = 0.05$

»  $n_1 = 8$

»  $n_2 = 10$

» Critical Region:  $\mu \leq 17$  (From Table)

» Computation Steps:

→ Arranging observations in ascending order and assigning ranks from 1 to 18.

$H_0$	$H_1$	Compute
	$\mu_1 < \mu_2$	$u_1$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$u_2$
	$\mu_1 \neq \mu_2$	$u$







DATA	RANKS	BRAND
0.6	1	B
1.6	2	B
1.9	3	B
2.1	4	A
2.2	5	B
2.5	6	B
3.1	7	B
3.3	8	A
3.7	9	A
4.0	10.5	A
4.0	10.5	B
4.1	12	B
4.8	13	A
5.4	14.5	A
5.4	14.5	B
6.1	16	A
6.2	17	B
6.3	18	A

- $w_1 = 4+8+9+10.5+13+14.5+16+18 = 93$
- $w_2 = 1+2+3+5+6+7+10.5+12+14.5+17 = 78$
- Therefore,  $u_1 = 93 - ((8*9)/2) = 57$
- $u_2 = 78 - ((10*11)/2) = 23$
- $\text{Min}(u_1, u_2) = 23$  ( not  $\leq 17$ )

Decision: Don't reject Null Hypothesis  $H_0$  and conclude that there is no significant difference in median nicotine contents of two brands of cigarettes at 0.05 significance level.



# Example (one tail)

In a genetic inheritance study discussed by Margolin[1988], samples of individuals from several ethnic groups were taken. Blood samples were collected from each individual and several variables measured.

Here, we want to test the hypothesis that the median of blood samples for Native Americans is the same as that for Caucasians against the alternative hypothesis that the median of the blood samples for Native Americans is **less than** that for Caucasians.



A	B
8.50	8.27
9.48	8.20
8.65	8.25
8.16	8.14
8.83	9.00
7.76	8.10
8.63	7.20
	8.32
	7.70

$$n_1 = 8$$

$$n_2 = 10$$

		Rank
7.20	B	1
7.70	B	2
7.76	A	3
8.10	B	4
8.14	B	5
8.16	A	6
8.20	B	7
8.25	B	8
8.27	B	9
8.32	B	10
8.50	A	11
8.63	A	12
8.65	A	13
8.83	A	14
9.00	B	15
9.48	A	16



- »  $H_0 : \mu_1 = \mu_2$
- »  $H_1 : \mu_1 < \mu_2$
- »  $\alpha = 0.05$
- » Critical Region:  $\mu \leq 12$  (from Table)
- »  $w_1 = 3 + 6 + 11 + 12 + 13 + 14 + 16 = 75$
- »  $w_2 = 1 + 2 + 4 + 5 + 7 + 8 + 9 + 10 + 15 = 61$

- »  $u_1 = 75 - \frac{(7)(8)}{2} = 47$
- »  $u_2 = 61 - \frac{(9)(10)}{2} = 16$

$H_0$	$H_1$	Compute
	$\mu_1 < \mu_2$	$u_1$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$u_2$
	$\mu_1 \neq \mu_2$	$u$

- » Conclusion: Do not reject the null hypothesis  $H_0$  and conclude that there is no significant difference in the median of the two brands.

