Wilcoxon Rank-Sum Test, also known as the Mann-Whitney test

- Rank the data. That is, replace the data values by their ranks, from smallest to largest. For example, the pH samples are:

Loc									
1	8.53	8.52	8.01	7.99	7.93	7.89	7.85	7.82	7.80
2	7.85	7.73	7.58	7.40	7.35	7.30	7.27	7.27	7.23

  are replaced by the ranks

Loc									
1	18	17	16	15	14	13	11.5	10	9
2	11.5	8	7	6	5	4	3	2	1

  The tied values (7.85, 7.85) would have had ranks 11 and 12 were they slightly different. In the case of ties, assign the "mid-rank" [here (11+12)/2] to both values.

- Calculate W, which is the sum of the ranks in the first group. In this case, $W = 123.5$.

- In small samples we compare W to the distribution of values of W under all possible allocations of the ranks to the two samples.

- In larger samples we use a normal approximation to this distribution.

- Let n_1 be the number of observations in the first group, n_2 the number in the second group, and $N = n_1 + n_2$. Here $n_1 = n_2 = 9$ and $N = 18$.

- Under the null hypothesis that the two distributions are the same, the mean and variance of W are
$$\mu_W = n_1(N+1)/2$$
and
$$\sigma_W^2 = n_1 n_2 (N+1)/12$$

- the observed value of the test statistic is

$$Z_{obs} = \frac{W - \mu_W}{\sigma_W} = \frac{(123.5 - 9 \times 19/2)}{\sqrt{9 \times 9 \times 19/12}} = 3.355$$

- As with the permutation test, the null hypothesis is that the population distributions are the same, and the two sided alternative is that the distributions are different.

- The hypotheses can also be written in terms of means. Where μ_1 and μ_2 are the means of the two populations, the null hypothesis is $H_0 : \mu_1 = \mu_2$. The possible alternatives and p-values are:

H_A	p-value
$\mu_1 \neq \mu_2$	$2P(Z > \lvert Z_{obs} \rvert)$
$\mu_1 > \mu_2$	$P(Z > Z_{obs})$
$\mu_1 < \mu_2$	$P(Z < Z_{obs})$

- For example, with the two sided alternative, the p-value is

$$2P(Z > |3.355|) = 2(.0004) = .0008$$

Comparison of Wilcoxon, Permutation and t-tests

- What p-value did the permutation test give in this case?

  The test statistic for the permutation test was $|\bar{X}_1 - \bar{X}_2| = 0.595556$.

  With a two sided alternative, there were $\begin{pmatrix} 18 \\ 9 \end{pmatrix} =$ 48620 possible arrangements of the data into two groups of size 9, of which 8 gave at least as much evidence against H_0 as did the value 3.335. Hence the p-value for the permutation test was
  $\mathrm{P} = 8/48620 = .0002$.

- The p-value for the pooled t-test was 0.0002.
- In this case, the Wilcoxon, permutation and t-tests all show very strong evidence against the null hypothesis of equal distributions.
- Typically, the p-values for the permutation and Wilcoxon test will be fairly close if the sample size is moderately large.

- If the data are approximately normally distributed and the assumption of equal variances holds, the p-value for the pooled t-test will be fairly close to those for the Wilcoxon and permutation tests.

Example 2: $n_1 = n_2 = 3$ with data

Treatment	175	250	260
Control	255	275	300

- The treatment ranks are 1,2,4.
- The control ranks are 3,5,6.
- The sum of the treatment ranks is $W = 7$.
- In this case is it possible to consider all possible permutations of the ranks between the two samples

1	2	3	4	5	6	W
1	1	1	2	2	2	6
1	1	2	1	2	2	$7 = W_{obs}$
1	1	2	2	1	2	8
1	1	2	2	2	1	9
1	2	1	2	2	1	10
1	2	1	2	1	2	9
1	2	1	1	2	2	8
1	2	2	1	1	2	10
1	2	2	2	1	1	12
1	2	2	1	2	1	11
2	1	1	1	2	2	9
2	1	1	2	1	2	10
2	1	1	2	2	1	11
2	1	2	1	1	2	11
2	1	2	1	2	1	12
2	1	2	2	1	1	13
2	2	1	1	1	2	12
2	2	1	1	2	1	13
2	2	1	2	1	1	14
2	2	2	1	1	1	15

- The distribution of W under H_o is summarized in the table below

W	6	7	8	9	10	11	12	13	14	15
Prob	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$

- The probability of getting a value for W as small or smaller than $W_{obs} = 7$ is $2/20$, and the probability of getting a value as extreme or more extreme is $P = 4/20 = .2$
- Under the null hypothesis, the mean of W is $3 \times 7/2 = 10.5$ and the variance of W is $3 \times 3 \times 7/12 = 5.25$.
- Using the normal approximation to the null distribution of W, the observed test statistic is $Z_{obs} = (7 - 10.5)/\sqrt{5.25} = -1.53$.
- With the two sided alternative, the p-value is

  $$2P(Z > |-1.53|) = 2P(Z > 1.53) = 2(.063) = .126$$

- Even with such small samples, the two approaches give quite similar answers.