3.7 Permutation and Randomization Tests

Permutation or **Randomization** tests are useful in statistical hypothesis testing as they are distribution free, and do not depend on any large sample approximations.

The methods can be applied to any statistical testing procedure - we will study a particular case, the non-parametric two sample comparison.

Basic Principle : For the hypothesis

$$H_0 \; : \; \text{No difference between populations}$$

the implication of H_0 is that **group labels are unimportant**. That is, we can permute the group labels without changing the value of the chose test statistic to any great extent.

Example: $n_1 = 3, n_2 = 2$.

	Sample 1			Sample 2	
Observed Data	0.1	3.4	1.6	2.4	1.1
Permuted Data	0.1	1.1	3.4	1.6	2.4
	0.1	1.1	2.4	1.6	3.4
	1.6	2.4	3.4	0.1	1.1
	⋮	⋮	⋮	⋮	⋮

We always obtain a 3/2 split: we choose (without replacement)

- 3 at random for the first group
- 2 at random for the second group

There are

$$\binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \times 4}{1 \times 2} = 10$$

possible splits. In general, there are

$$\binom{n_1 + n_2}{n_1} = \binom{n_1 + n_2}{n_2}$$

possible splits.

If H_0 is true, all such splits are equally likely to occur in the observed data.

Thus if we compute the test statistic for each split, then the resulting distribution of statistics is **precisely** the distribution of the test statistic under H_0.

This distribution forms the basis of the statistical test.

SEE HANDOUT

NON-PARAMETRIC STATISTICS

THE ROLE OF RANDOMIZATION/PERMUTATION TESTS

Randomization or **Permutation** procedures are useful for computing **exact** null distributions for non-parametric test statistics when sample sizes are small.

We focus first on two sample comparisons; suppose that two data samples $x_1 \ldots, x_{n_1}$ and $y_1 \ldots, y_{n_2}$ (where $n_1 \geq n_2$) have been obtained, and we wish to carry out a comparison of the two populations from which the samples are drawn. The Wilcoxon test statistic, W, is the sum of the ranks for the second sample. The permutation test proceeds as follows:

1. Let $n = n_1 + n_2$. Assuming that there are no ties, the pooled and ranked samples will have ranks

$$1 \quad 2 \quad 3 \quad \ldots \quad n$$

2. The test statistic is $W = R_2$, the rank sum for sample two items. For the observed data, W will be the sum of n_2 of the ranks given in the list above.

3. If the null hypothesis

$$H_0 \ : \ \text{No difference between population 1 and population 2}$$

   were **true**, then we would expect **no pattern** in the arrangements of the group labels when sorted into ascending order. That is, the sorted data would give rise a **random** assortment of group 1 and group 2 labels.

4. To obtain the exact distribution of W under H_0 (which is what we require for the assessment of statistical significance), we could compute W for all possible permutations of the group labels, and then form the probability distribution of the values of W. We call this the **permutation null distribution**.

5. But W is a rank sum, so we can compute the permutation null distribution simply by tabulating **all possible subsets** of size n_2 of the set of ranks $\{1, 2, 3, \ldots, n\}$.

6. There are

$$\binom{n}{n_2} = \frac{n!}{n_1!\, n_2!} = N$$

say possible subsets of size n_2. For example, for $n = 6$ and $n_2 = 2$, the number of subsets of size n_2 is

$$\binom{8}{2} = \frac{8!}{6!\, 2!} = 28$$

However, the number of subsets increases dramatically as n increases; for $n_1 = n_2 = 10$, so that $n = 20$, the number of subsets of size n_2 is

$$\binom{20}{10} = \frac{20!}{10!\, 10!} = 184756$$

7. The exact rejection region and p-value are computed from the permutation null distribution. Let $W_i, i = 1, \ldots, N$ denote the value of the Wilcoxon statistic for the N possible subsets of the ranks of size n_2. The probability that the test statistic, W, is less than or equal to w is

$$\Pr[W \leq w] = \frac{\text{Number of } W_i \leq w}{N}$$

We seek the values of w that give the appropriate rejection region, $\mathcal{R}$, so that

$$\Pr[W \in \mathcal{R}] = \frac{\text{Number of } W_i \in \mathcal{R}}{N} = \alpha$$

It may not be possible to find critical values, and define $\mathcal{R}$, so that this probability is **exactly** α as the distribution of W is **discrete**.

Simple Example

Suppose $n_1 = 7$ and $n_2 = 3$. There are

$$\binom{10}{3} = \frac{10!}{7!\,3!} = 120$$

subsets of the ranks $\{1, 2, 3, \ldots, 10\}$ of size 3. The subsets are listed below, together with the rank sums.

Ranks			W	Ranks			W	Ranks			W	Ranks			W
1	2	3	6	1	7	8	16	2	7	10	19	4	6	7	17
1	2	4	7	1	7	9	17	2	8	9	19	4	6	8	18
1	2	5	8	1	7	10	18	2	8	10	20	4	6	9	19
1	2	6	9	1	8	9	18	2	9	10	21	4	6	10	20
1	2	7	10	1	8	10	19	3	4	5	12	4	7	8	19
1	2	8	11	1	9	10	20	3	4	6	13	4	7	9	20
1	2	9	12	2	3	4	9	3	4	7	14	4	7	10	21
1	2	10	13	2	3	5	10	3	4	8	15	4	8	9	21
1	3	4	8	2	3	6	11	3	4	9	16	4	8	10	22
1	3	5	9	2	3	7	12	3	4	10	17	4	9	10	23
1	3	6	10	2	3	8	13	3	5	6	14	5	6	7	18
1	3	7	11	2	3	9	14	3	5	7	15	5	6	8	19
1	3	8	12	2	3	10	15	3	5	8	16	5	6	9	20
1	3	9	13	2	4	5	11	3	5	9	17	5	6	10	21
1	3	10	14	2	4	6	12	3	5	10	18	5	7	8	20
1	4	5	10	2	4	7	13	3	6	7	16	5	7	9	21
1	4	6	11	2	4	8	14	3	6	8	17	5	7	10	22
1	4	7	12	2	4	9	15	3	6	9	18	5	8	9	22
1	4	8	13	2	4	10	16	3	6	10	19	5	8	10	23
1	4	9	14	2	5	6	13	3	7	8	18	5	9	10	24
1	4	10	15	2	5	7	14	3	7	9	19	6	7	8	21
1	5	6	12	2	5	8	15	3	7	10	20	6	7	9	22
1	5	7	13	2	5	9	16	3	8	9	20	6	7	10	23
1	5	8	14	2	5	10	17	3	8	10	21	6	8	9	23
1	5	9	15	2	6	7	15	3	9	10	22	6	8	10	24
1	5	10	16	2	6	8	16	4	5	6	15	6	9	10	25
1	6	7	14	2	6	9	17	4	5	7	16	7	8	9	24
1	6	8	15	2	6	10	18	4	5	8	17	7	8	10	25
1	6	9	16	2	7	8	17	4	5	9	18	7	9	10	26
1	6	10	17	2	7	9	18	4	5	10	19	8	9	10	27

There are 22 possible rank sums, $\{6, 7, 8, \ldots, 25, 26, 27\}$; the number of times each is observed is displayed in the table below, with the corresponding probabilities and cumulative probabilities.

W	6	7	8	9	10	11	12	13	14	15	16
Frequency	1	1	2	3	4	5	7	8	9	10	10
Prob.	0.008	0.008	0.017	0.025	0.033	0.042	0.058	0.067	0.075	0.083	0.083
Cumulative Prob.	0.008	0.017	0.033	0.058	0.092	0.133	0.192	0.258	0.333	0.417	0.500

W	17	18	**19**	20	21	22	23	24	25	26	27
Frequency	10	10	**9**	8	7	5	4	3	2	1	1
Prob.	0.083	0.083	**0.075**	0.067	0.058	0.042	0.033	0.025	0.017	0.008	0.008
Cumulative Prob.	0.583	0.667	0.742	0.808	0.867	0.908	0.942	0.967	0.983	0.992	1.000

Thus, for example, the probability that $W = 19$ is 0.075, with a frequency of 9 out of 120. From this table, we deduce that

$$\Pr[8 \leq W \leq 25] = 0.983 - 0.017 = 0.966$$

implying that the two-sided rejection region for $\alpha = 0.05$ is the set $\mathcal{R} = \{6, 7, 26, 27\}$.

Placenta Permeability Data

Using the placenta permeability data from Assignment 3, the data and ranks for are displayed below:

Group	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
Obs.	0.73	0.80	0.83	1.04	1.38	1.45	1.46	1.64	1.89	1.91	0.74	0.88	0.9	1.15	1.21
Rank	1	3	4	7	10	11	12	13	14	15	2	5	6	8	9

Thus the Wilcoxon statistic is

$$W = R_2 = 2 + 5 + 6 + 8 + 9 = 30$$

Now, here $n_1 = 10$ and $n_2 = 5$. There are

$$\binom{15}{5} = \frac{15!}{10!\,5!} = 3003$$

subsets of the ranks $\{1, 2, 3, \ldots, 15\}$ of size 5.

In the permutation null distribution, the possible values of W are $\{15, 16, \ldots, 64, 65\}$; the probabilities are given below.

W	15	16	17	18	19	20	21	22	23	24	25	26	27
Frequency	1	1	2	3	5	7	10	13	18	23	30	36	45
Prob.	0.000	0.000	0.001	0.001	0.002	0.002	0.003	0.004	0.006	0.008	0.010	0.012	0.015
Cumulative Prob.	0.000	0.001	0.001	0.002	0.004	0.006	0.010	0.014	0.020	0.028	0.038	0.050	0.065
W	28	29	30	31	32	33	34	35	36	37	38	39	40
Frequency	53	63	72	83	92	103	111	121	127	134	137	141	141
Prob.	0.018	0.021	0.024	0.028	0.031	0.034	0.037	0.040	0.042	0.045	0.046	0.047	0.047
Cumulative Prob.	0.082	0.103	0.127	0.155	0.185	0.220	0.257	0.297	0.339	0.384	0.430	0.477	0.523
W	41	42	43	44	45	46	47	48	49	50	51	52	53
Frequency	141	137	134	127	121	111	103	92	83	72	63	53	45
Prob.	0.047	0.046	0.045	0.042	0.040	0.037	0.034	0.031	0.028	0.024	0.021	0.018	0.015
Cumulative Prob.	0.570	0.616	0.661	0.703	0.743	0.780	0.815	0.845	0.873	0.897	0.918	0.935	0.950
W	54	55	56	57	58	59	60	61	62	63	64	65	
Frequency	36	30	23	18	13	10	7	5	3	2	1	1	
Prob.	0.012	0.010	0.008	0.006	0.004	0.003	0.002	0.002	0.001	0.001	0.000	0.000	
Cumulative Prob.	0.962	0.972	0.980	0.986	0.990	0.994	0.996	0.998	0.999	0.999	1.000	1.000	

By inspection of the table, we see that

$$\Pr[25 \leq W \leq 55] = 0.972 - 0.028 = 0.944$$

and

$$\Pr[24 \leq W \leq 56] = 0.980 - 0.020 = 0.960$$

Thus for a symmetric two-sided interval which contains at most probability 0.95, we take the interval

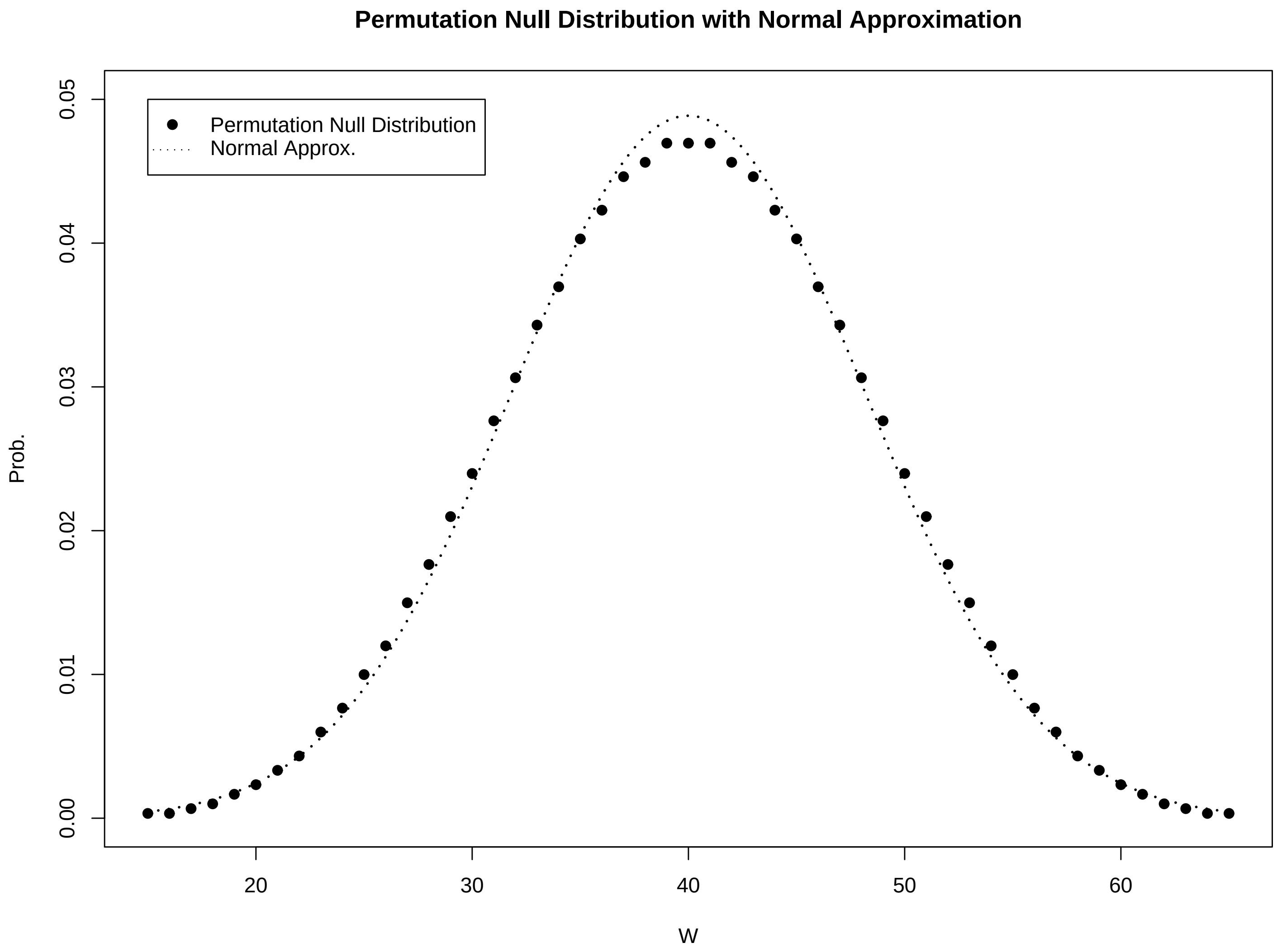
$$\{25, 26, \ldots, 54, 55\}$$

and hence define the rejection region

$$\mathcal{R} = \{15, 16, \ldots, 23, 24, 56, 57, \ldots, 64, 65\}$$

Note that this choice of rejection region ensures that there is at least probability 0.025 in each tail.

The permutation null distribution of W is displayed below.


Permutation Null Distribution with Normal Approximation
Permutation Null Distribution
Normal Approx.
Prob.
0.00
0.01
0.02
0.03
0.04
0.05
20
30
40
50
60
W

The normal approximation is given by

$$W \mathrel{\dot\sim} \text{Normal}\left(\frac{n_2(n_1+n_2+1)}{2}, \frac{n_1 n_2(n_1+n_2+1)}{12}\right)$$