Part III

Non-Parametric Statistics

Non-Parametric Statistics

All of the previous statistical analysis methods studied (t-tests, ANOVA, Regression, General Linear Modelling) have depended heavily on **distributional assumptions**.

i.e. we assume that the data are **Normally distributed**.

We now seek statistical procedures that do not rely on this strong assumption. We term these methods

NON-PARAMETRIC

or

DISTRIBUTION-FREE

They substitute **large sample approximations** for the distributional assumptions.

3.1 Distribution-free tests for Categorical Data

Categorical data are data in which experimental units are allocated to one of a number of categories according to their characteristics. The categories are defined by one or more factors

Examples:

- Female/Male - two categories
- Smoker/Former Smoker/Non Smoker - three categories.

Doll and Hill Data

Table 13.11. Smokers and non-smokers among male cancer patients and controls (Doll and Hill 1950)

	Smokers	Non-smokers	Total
Lung cancer	647	2	649
Controls	622	27	649

Juvenile Delinquency and Spectacle-Wearing

Table 10.14 Spectacle wearing among juvenile delinquen and non-delinquents who failed a vision test (Weindling *al* 1986)

		Juvenile delinquents	Non delinquents	Total
Spectacle wearers	Yes	1	5	6
	No	8	2	10
	Tota	9	7	16

The data are **counts** of experimental units that fall into each category. Suppose

1. There are n experimental units in the study
2. There are k categories
3. The probabilities of the k outcomes are $p_1, \ldots, p_k$, where

$$p_1 + \cdots + p_k = 1$$

4. The experimental units are **independent**
5. The counts in the k categories are $n_1, \ldots, n_k$, where

$$n_1 + \cdots + n_k = n$$

The experimental design is termed a **Multinomial Experiment**

Note: The categories can arise as combinations of factor levels; we can have

- **one-way** classification (categories of a single factor, A)
- **two-way** classification (categories defined by combinations of levels of two factors, A and B)

and so on. The counts table is often called a **contingency table** and the entries in the table are called **cells**.

The idea can be extended to larger numbers of factors $(A, B, C, \ldots)$ to produce a multi-way table. We will focus on at most two-way tables, with r rows and c columns, yielding an $r \times c$ table.

What kinds of tests can be carried out for such data ?

1. Tests about $p_1, \ldots, p_k$
   - $H_0 : p_1 = \cdots = p_k = 1/k$
   - $H_0 : p_1, \ldots, p_k$ determined by some parametric distribution (Normal, Poisson etc.)
2. Tests about the factors A and B
   - are A and B dependent ?
   - i.e. does classification by A influence classification by B.

Chi-Squared Test

For one-way tables: suppose that a null hypothesis **completely specifies** $p_1, \ldots, p_k$, that is, we have H_0 of the form

$$H_0 \ : \ p_1 = p_1^{(0)}, \ldots, p_k = p_k^{(0)}$$

where $p_1^{(0)}, \ldots, p_k^{(0)}$ are fixed probabilities. For example, for $k = 3$,

$$H_0 \ : \ p_1 = p_2 = p_3 = 1/3$$

or

$$H_0 \ : \ p_1 = 1/2, p_2 = p_3 = 1/4$$

To test this hypothesis against the general alternative hypothesis

$$H_a \ : \ H_0 \text{ not true.}$$

we use the test statistic

$$X^2 = \sum_{i=1}^{k} \frac{\left(n_i - np_i^{(0)}\right)^2}{np_i^{(0)}}$$

If H_0 is true,

$$X^2 \dot{\sim} \text{ Chi-squared}(k-1).$$

that is, X^2 is approximately distributed as Chi-squared$(k-1)$.

In this formula

- n_i is the **observed** count in cell i
- $np_i^{(0)}$ is the **expected** count in cell i if H_0 is **true**.

Sometimes the formula is written

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed count, and E_i is the expected count.

If

$$X^2 > \text{Chisq}_\alpha(k-1)$$

then we reject H_0 at the α significance level, where $\text{Chisq}_\alpha(k-1)$ is the $1-\alpha$ (right-hand) tail critical value of the Chi-squared distribution with $k-1$ degrees of freedom.

This method can be extended in the one-way case to test distribution assumptions, that is, for example

$$H_0 \ : \ \text{Data Normally distributed}$$

or

$$H_0 \ : \ \text{Data Poisson distributed}$$

Unfortunately this facility is not available in SPSS; direct calculation is possible but involved.

For the **two-way** table, we can test the hypothesis

H_0 : Factor A and Factor B levels are assigned independently

that is, classification by factor A is independent of classification by factor B. We use the same test statistic that can be rewritten

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - \widehat{n}_{ij})^2}{\widehat{n}_{ij}}$$

where

$$\widehat{n}_{ij} = \frac{n_{i.} n_{.j}}{n} \qquad n_{i.} = \sum_{j=1}^{c} n_{ij} \qquad n_{.j} = \sum_{i=1}^{r} n_{ij}.$$

The terms $n_{i.}$ and $n_{.j}$ are the row and column totals for row i and column j respectively.

If H_0 is true

$$X^2 \dot{\sim} \text{Chi-squared}((r-1)(c-1))$$

i.e. the degrees of freedom quantity is $(r-1)(c-1)$. Otherwise the test proceeds as before: we check whether

$$X^2 > \text{Chisq}_\alpha((r-1)(c-1))$$

and if so, we reject H_0.

Example (DNA Sequence Data)

Counts of Nucleotides A,C,G,T in a genomic segment related to the breast cancer gene BRCA2.

Example (Eye and Hair Colour Data)

The assignment of hair and eye colour in a sample of humans

See handout.

Note: For the Chi-squared test to be valid, we need the expected cell counts

$$np_i^{(0)} \qquad i = 1, \ldots, k$$

or

$$\widehat{n}_{ij} \qquad i = 1, \ldots, r, j = 1, \ldots, c$$

to be sufficiently large. The convention is to require the expected value to be greater than **five**.

Note: If $r = c = 2$ we have a 2×2 table, and another **exact** test can be used which does not rely on the large sample approximation

Fisher's Exact Test

- another test for independence of assignment of the row and column factor levels
- test statistic and null distribution are complicated (based on the **hypergeometric distribution**)
- SPSS computes test statistic and p-value.

Example (Juvenile Delinquency and Spectacle Wearing)

Is there any association between the two factors ?

A : Spectacle Wearing (Yes/No)

B : Juvenile Delinquent (Yes/No)

		Delinquent		
		Yes	No	$n_{i.}$
Spectacles	Yes	1	5	6
	No	8	2	10
	$n_{.j}$	9	7	16

Example (Juvenile Delinquency and Spectacle Wearing)

Chi-squared Test:

$$X^2 = 6.112$$

Compare with Chi-squared$((r-1)(c-1))$ = Chi-squared(1); we have

$$\text{Chi-squared}_{0.05}(1) = 3.841$$

and a p-value of 0.013. Therefore we **reject** H_0.

Fisher's Exact Test: p-value is 0.035 (1-sided) or 0.024 (2-sided).

Thus we reject H_0 and we have evidence of association between the factors.

Case-Control Studies

A **case-control** study is an observational study where participants are selected for the study with regard to their **disease status**.

- a sample of **cases** (disease sufferers)
- a sample of **controls** (healthy patients)

We investigate the possible association between disease status and a factor that takes two levels. A 2×2 table of counts is formed for all combinations of disease status/factor level.

Example (BCG Vaccination and Leprosy)

Disease Status : Leprosy Sufferer (Yes/No)

Factor : Vaccination Scar (Yes/No)

		Disease Status: Case	Disease Status: Control	
		Yes	No	$n_{i.}$
Scar	Yes	101	554	655
	No	159	446	605
	$n_{.j}$	260	1000	1260

Is there an association ? Does vaccination induce leprosy ?

The Chi-squared test is potentially not valid here because of the design. An alternative test statistic is based on the **odds ratio**

$$\text{O.R.} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \widehat{\psi}$$

say. The test statistic is

$$Z = \frac{\log \widehat{\psi}}{\text{s.e.}(\log \widehat{\psi})}$$

where

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

That is,

$$Z = \frac{\log n_{11} + \log n_{22} - \log n_{12} - \log n_{21}}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}}$$

Under

$$H_0 \ : \text{ No association between factor and disease status}$$

it follows that

$$Z \dot{\sim} N(0, 1)$$

Here log means ln or **natural log**.

Example (BCG Vaccination and Leprosy)

$$n_{11} = 101, n_{12} = 554, n_{21} = 159, n_{22} = 446$$

Therefore

$$\widehat{\psi} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = 0.511 \qquad \log \widehat{\psi} = -0.671$$

and

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} = 0.142$$

so

$$Z = \frac{-0.671}{0.142} = -4.717$$

For a text at $\alpha = 0.05$, the two-sided critical value is ± 1.96, so we

Reject H_0.

Example (Smoking and Lung Cancer)

$$n_{11} = 647, n_{12} = 622, n_{21} = 2, n_{22} = 27$$

Therefore

$$\log \widehat{\psi} = \log \frac{647 \times 27}{2 \times 622} = 2.642$$

and

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{647} + \frac{1}{2} + \frac{1}{622} + \frac{1}{27}} = 0.735$$

so

$$Z = \frac{2.642}{0.735} = 3.590$$

For a text at $\alpha = 0.05$, the two-sided critical value is ± 1.96, so we

Reject H_0

and report evidence for association.

3.2 Single Population Tests

We seek non-parametric or distribution-free tests for hypotheses relating to single samples, the equivalents of one-sample Z- or T-tests, which rely on the **normality** of the samples.

Normally these tests are formulated in terms of **ranks** of the data to give

Rank Tests

For example, if the data are

$$0.24 \quad 3.16 \quad 1.97 \quad 2.10 \quad 0.92$$

we sort them into **ascending** order, and assign ranks in order

	0.24	0.92	1.97	2.10	3.16
Rank	1	2	3	4	5

The tests depend on the behaviour of statistics computed in terms of the ranks, and rely on a **large sample** justification.

Rather than test the **mean**, we test the **median**, x_{MED}, where

$$\Pr[\text{Observation } \leq x_{\text{MED}}] = \frac{1}{2}$$

i.e. the halfway point of the distribution.

The **sample median** is the halfway point of the sorted sample.

Let η denote the population median. We wish to test, for example,

$$H_0 : \eta = \eta_0$$

See Handout

3.3 Comparing Two Populations : Independent Samples

We seek a non-parametric equivalent to the two-sample t-test.

Instead of testing population **means**,

$$H_0 \ : \ \mu_1 = \mu_2$$

we test population **medians**

$$H_0 \ : \ \eta_1 = \eta_2$$

One- and Two- sample tests

- In the **one sample** case we use the

  SIGN TEST

  to test hypotheses about η

- In the **two sample** case we use the

  WILCOXON RANK SUM or **MANN-WHITNEY U** test.

See Handout

Note: For the MWW test

- **Textbook convention :** Label the samples so that $n_1 > n_2$ (i.e. sample 1 is the one with the larger sample size)
- **SPSS convention :** Label the samples such that

$$x_{\text{MED}_1} < x_{\text{MED}_2}$$

(i.e. sample 1 is the one with the smaller median) and only test

$$H_0 \ : \ \eta_1 = \eta_2$$

Other two sample tests are available:

- Kolmogorov-Smirnov Test
- Moses Extreme Reactions Test
- Wald-Wolfowitz Runs Test

None make distributional assumptions, all perform best when the sample size is large.

3.4 Comparing Two Dependent Samples

Suppose we have repeat measurements on the same experimental units.

In this case, the **within-subject** data are **dependent**; we have pairing of observations.

We can use the

Wilcoxon Signed Rank Test

See Handout

3.4 Comparing Three or More Populations

We now seek non-parametric equivalents to ANOVA useful for different designs. We study tests for

(a) the **Completely Randomized Design** (CRD)

(b) the **Randomized Block Design** (RBD)

For (a) we use the

Kruskal-Wallis Test

and for (b) we use the

Friedman Test.

See Handout

Summary of the Non-Parametric Tests

- **Chi-Squared Test :** Goodness of Fit/independence in contingency tables
- **Sign Test :** One Sample (equivalent of one sample t-test)
- **Mann-Whitney-Wilcoxon :** Two Sample (equivalent of two sample t-test)
- **Wilcoxon Signed Rank :** Paired Data
- **Kruskal-Wallis :** one-way layout, multigroup comparison - equivalent of ANOVA for CRD.
- **Friedman :** two-way blocked layout, equivalent of two-way ANOVA for RBD.

Pros:

- No distributional assumptions
- Applicable for most sorts of data
- Large sample approximations make them easy to implement

Cons:

- Low power compared to parametric tests (i.e. often do not reject H_0 when they should - prone to Type II Error)
- Small sample null distributions difficult to compute.

3.6 Rank Correlation

To measure the association between two variables, we previously used the *correlation coefficient*, r; for data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$,

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

where

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \quad SS_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 \quad SS_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

r is a measure of the linear association between X and Y

Pearson Product Moment Coefficient of Correlation

A more general measure of association is the

Spearman Rank Correlation Coefficient

We compute this as follows:

1. For each sample separately, compute the **ranks** of the data, denote the ranks for the data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ by $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ respectively.

2. Compute

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}SS_{vv}}}$$

   ie the Pearson correlation between the ranks.

r_S is the **Spearman Correlation**.

Notes:

1. If there are no ties in the data

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

   where $d_i = u_i - v_i$.

2. r_S is potentially a measure of the **non-linear** association between X and Y.

   The calculation can be applied directly to rank data i.e. $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ can be preference ranks given by two observers.

Tests for r_S

To test

$$H_0 \ : \ \rho = 0$$

vs

$$\begin{array}{lll} (1) \quad H_a & : & \rho > 0 \\ (2) \quad H_a & : & \rho < 0 \\ (3) \quad H_a & : & \rho \neq 0 \end{array}$$

We may use r_S as a test statistic. The distribution of r_S under H_0 is tabulated on p 864 of McClave and Sincich.

If Spearman_α is the α tail quantile of the null distribution, we have the following rejection regions:

$$
\begin{aligned}
(1) \quad &: \quad \text{Reject } H_0 \text{ if } r_S > \text{Spearman}_\alpha \\
(2) \quad &: \quad \text{Reject } H_0 \text{ if } r_S < -\text{Spearman}_\alpha \\
(3) \quad &: \quad \text{Reject } H_0 \text{ if } |r_S| > \text{Spearman}_{\alpha/2}
\end{aligned}
$$

The Role of Randomization and 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 : 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

$$\Pr[8 \leq W \leq 25] = 0.983 - 0.033 = 0.950$$

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

Placenta Permeability Data

Example (Placenta Permeability Data)

Measurements of placenta permeability are made on two groups of subjects.

The data and their 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

Placenta Permeability Data

Example

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.

Placenta Permeability Data

Example

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

Placenta Permeability Data

Example

By inspection of the table, we see that

$$\Pr[25 \leq W \leq 55] = 0.972 - 0.038 = 0.934$$

and

$$\Pr[24 \leq W \leq 56] = 0.980 - 0.028 = 0.952$$

Placenta Permeability Data

Example

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

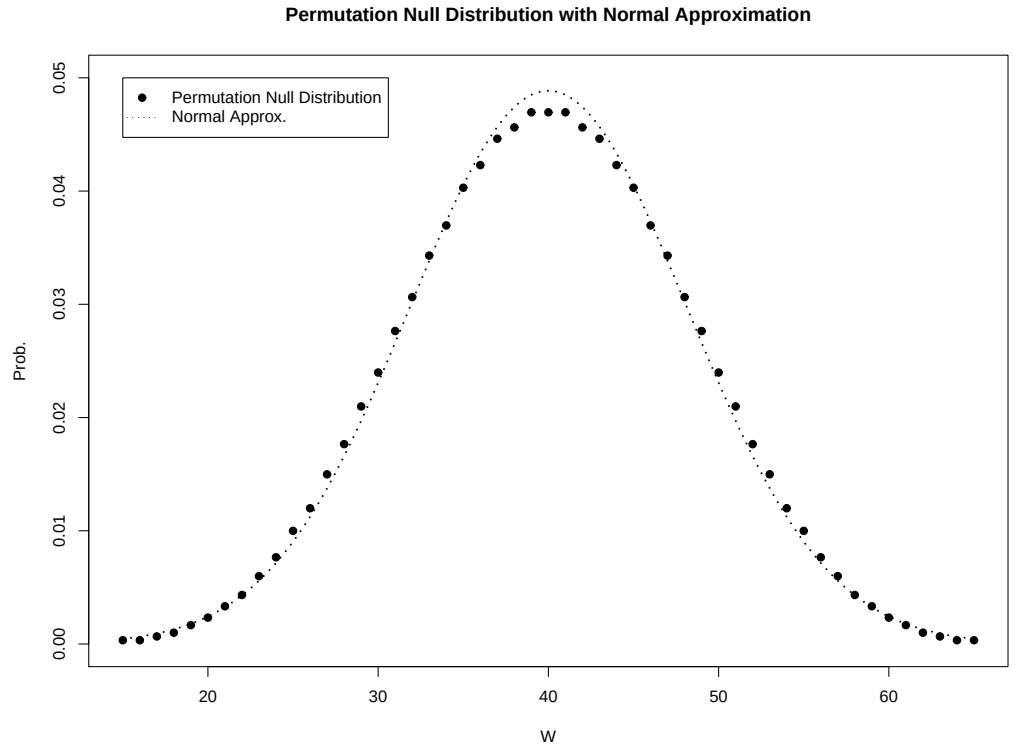
$$\mathcal{R} = \{16, 17, \ldots, 23, 24, 56, 57, \ldots, 64, 65\}$$

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

Thus in this example we **do not reject** the hypothesis of equal medians.

Normal Approximation

The permutation null distribution of W is displayed below.



The normal approximation is given by

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