

Chapter 3

Non-Parametric Statistics

In this section

- Textbook Chapters 13 & 14
- Chi-squared goodness of fit tests
- Non-parametric tests based on ranks
- One sample tests
- Two sample tests
- Permutation procedures

Non-Parametric Statistics

The previous statistical analysis methods studied (such as t-tests, ANOVA, Regression, Linear Modelling) have depended heavily on

distributional assumptions

specifically, 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

approaches.

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

- Treatment A, Treatment B - two categories
- Smoker/Former Smoker/Non Smoker - three categories.

Distribution-free tests for Categorical Data

In MATH 203: for 2×2 tables

- Two sample comparison of **proportions**
- Uses large sample approximation

See textbook, section 9.4.

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

Count data

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

- There are n experimental units in the study
- There are k categories
- The probabilities of the k outcomes are $p_1, \dots, p_k$, where

$$p_1 + \dots + p_k = 1$$

- The experimental units are **independent**
- The counts in the k categories are $n_1, \dots, n_k$, where

$$n_1 + \dots + n_k = n$$

Count data

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 (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**.

Count data

The idea can be extended to larger numbers of factors

$$A, B, C, \dots$$

to produce a **multi-way** table.

We will focus on two-way tables, with r rows and c columns, yielding an $r \times c$ table.

Recap: The Binomial Distribution

Suppose random variable X can take values on the set

$$\{0, 1, \dots, n\}$$

for some whole number $n \geq 1$. The **Binomial distribution** is a probability distribution for which

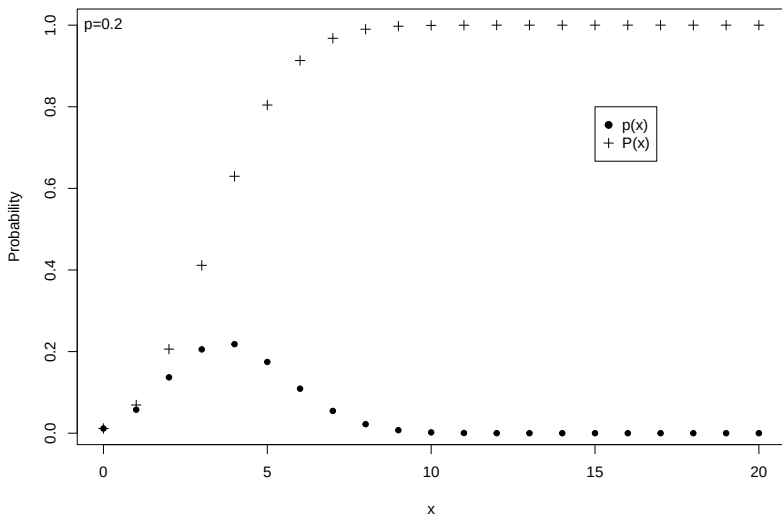
$$p(x) = P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x} \quad 0 \leq x \leq n$$

for some probability p .

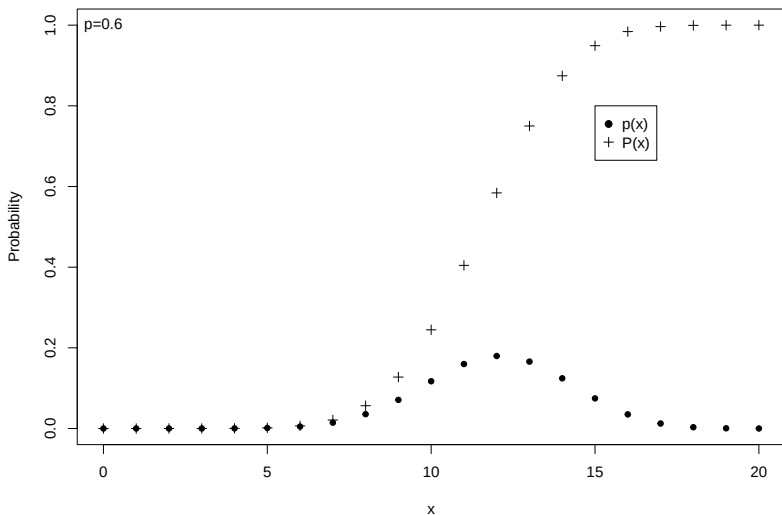
The **cumulative probability function**, $P(x)$, is defined by

$$P(x) = P[X \leq x] = \sum_{t=0}^x \binom{n}{t} p^t (1 - p)^{n-t}$$

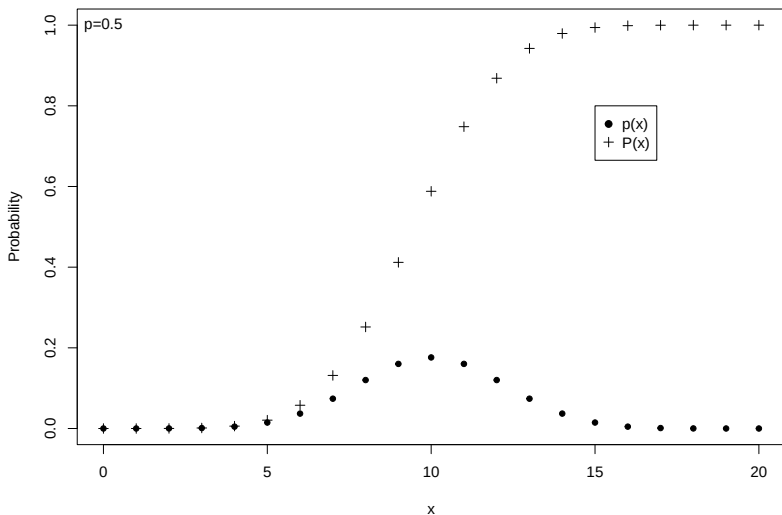
Recap: The Binomial Distribution



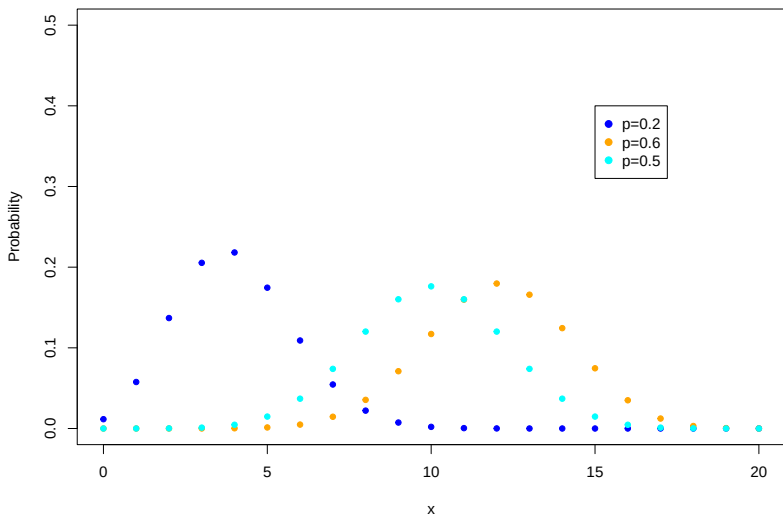
Recap: The Binomial Distribution



Recap: The Binomial Distribution



Recap: The Binomial Distribution



Tests for Binomial/Multinomial Data

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

1. Tests about $p_1, \dots, p_k$

- $H_0 : p_1 = \dots = p_k = 1/k$
- $H_0 : p_1, \dots, p_k$ determined by distribution (eg Normal)

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 the null hypothesis **completely specifies** $p_1, \dots, p_k$, that is, we have H_0 of the form

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

where $p_1^{(0)}, \dots, p_k^{(0)}$ are specified probabilities.

For example, for $k = 3$,

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

or

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

Chi-Squared Test for one-way tables

To test this hypothesis against the alternative hypothesis

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

we use the test statistic

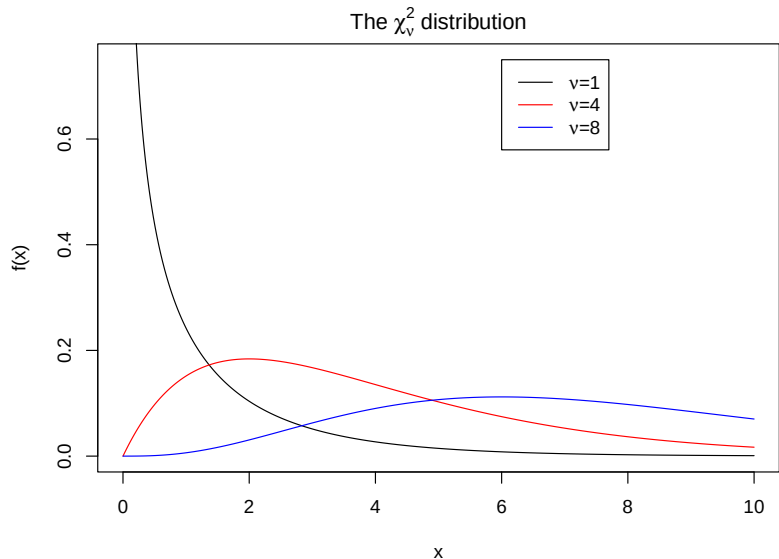
$$X^2 = \sum_{i=1}^k \frac{(n_i - np_i^{(0)})^2}{np_i^{(0)}}$$

If H_0 is true,

$$X^2 \rightsquigarrow \chi_{k-1}^2.$$

That is, X^2 is **approximately distributed** as a **chi-squared** distribution with $k - 1$ degrees of freedom.

Chi-Squared Test for one-way tables



Chi-Squared Test for one-way tables

Table: Upper a points of the χ^2_ν distribution

ν	a				ν	a			
	0.100	0.050	0.010	0.001		0.100	0.050	0.010	0.001
1	2.706	3.841	6.635	10.828	11	17.275	19.675	24.725	31.264
2	4.605	5.991	9.210	13.816	12	18.549	21.026	26.217	32.909
3	6.251	7.815	11.345	16.266	13	19.812	22.362	27.688	34.528
4	7.779	9.488	13.277	18.467	14	21.064	23.685	29.141	36.123
5	9.236	11.070	15.086	20.515	15	22.307	24.996	30.578	37.697
6	10.645	12.592	16.812	22.458	16	23.542	26.296	32.000	39.252
7	12.017	14.067	18.475	24.322	17	24.769	27.587	33.409	40.790
8	13.362	15.507	20.090	26.124	18	25.989	28.869	34.805	42.312
9	14.684	16.919	21.666	27.877	19	27.204	30.144	36.191	43.820
10	15.987	18.307	23.209	29.588	20	28.412	31.410	37.566	45.315

Chi-Squared Test for one-way tables

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

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

where O_i is the **observed** count, and E_i is the **expected** count assuming H_0 is true.

Chi-Squared Test for one-way tables

If

$$X^2 > \chi_{\alpha, k-1}^2$$

then we **reject** H_0 at the α significance level, where

$$\chi_{\alpha, k-1}^2$$

is the upper α point of the chi-squared distribution with $k - 1$ degrees of freedom.

Two-way tables

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 .

Two-way tables

We use the same test statistic that can be rewritten

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

where

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

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

Two-way tables

If H_0 is true

$$X^2 \approx \chi^2_{(r-1)(c-1)}$$

i.e. the degrees of freedom quantity is $(r - 1)(c - 1)$.

The test proceeds as before: we check whether

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

and if so, we **reject** H_0 .

Example: 2×3 contingency table

Patients suffering from a chronic condition are recruited to a randomized clinical trial with three treatment groups:

- Drug 1
- Drug 2
- Drug 3

The alleviation of the chronic condition was studied:

- Condition did not improve
- Condition did improve

Data:

Improved	Drug			Total
	1	2	3	
No	10	12	10	32
Yes	10	18	15	43
Total	20	30	25	75

Is there a **difference** between the improvement rates across drugs ?

Notation:

Improved	Drug			Total
	1	2	3	
No	n_{11}	n_{12}	n_{13}	$n_{1.}$
Yes	n_{21}	n_{22}	n_{23}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

Scientific question

Null Hypothesis:

H_0 : Improvement rate is **the same** across drugs.

If H_0 is true, there is **no difference** between the drugs

- the elements in the columns of the table are realizations of three independent Binomial sampling schemes with the **same** probability of success p .

The best estimate of the **common** improvement rate p is

$$\hat{p} = \frac{n_{2.}}{n} = \frac{43}{75} = 0.573$$

Scientific question

That is

$$n_{21} \sim \text{Binomial}(n_{.1}, p)$$

$$n_{22} \sim \text{Binomial}(n_{.2}, p)$$

$$n_{23} \sim \text{Binomial}(n_{.3}, p)$$

and thus, under H_0 , we use the fitted values determined by the expectations of these distributions

$$\hat{n}_{21} = n_{.1}\hat{p} = \frac{n_{.1}n_{2.}}{n}$$

$$\hat{n}_{22} = n_{.2}\hat{p} = \frac{n_{.2}n_{2.}}{n}$$

$$\hat{n}_{23} = n_{.3}\hat{p} = \frac{n_{.3}n_{2.}}{n}$$

Scientific question

We can complete the “fitted” or “expected” table by noting that we must have

$$n_{.1} = \hat{n}_{11} + \hat{n}_{21}$$

$$n_{.2} = \hat{n}_{12} + \hat{n}_{22}$$

$$n_{.3} = \hat{n}_{13} + \hat{n}_{23}$$

This is sufficient to allow us to complete the “fitted” table

Fitted table under H_0

Improved	Drug			Total
	1	2	3	
No	8.533	12.800	10.667	32
Yes	11.467	17.200	14.333	43
Total	20.000	30.000	25.000	75

so that the test statistic is

$$X_{obs}^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = 0.599$$

Completing the test

Under H_0 , we have that

$$X^2 \rightsquigarrow \chi^2_{(r-1)(c-1)}$$

Here $r = 2$, $c = 3$, so

$$X^2 \rightsquigarrow \chi^2_2$$

From the textbook tables Table VII, we find that

$$\chi^2_{0.05,2} = 5.991$$

so we **do not reject** the null hypothesis, and conclude that there is no evidence of a difference between the improvement rates for the three drugs.

Example: Nose bleeds and Alcohol

Example: Nose bleeds and alcohol consumption

140 patients suffering from nose bleeds, and 113 controls were interviewed about their alcohol consumption, specifically whether they were

1. Non-drinkers
2. Occasional drinkers
3. Regular drinkers

Scientific question

Is there a **difference** between the proportions of different consumption groups in nose-bleed cases and controls ?

Example: Nose bleeds and Alcohol

Improved	Group			Total
	1	2	3	
Nose-bleed	47	30	63	140
Control	39	40	34	113
Total	86	70	97	253

Example: Nose bleeds and Alcohol

Using the `chisq.test` function:

```
> NoseBleeds<-matrix(c(47,30,63,39,40,3), nrow=2,byrow=T)
> chisq.test(NoseBleeds)

+
+ Pearson's Chi-squared test
+
+ data:  NoseBleeds
+ X-squared = 44.61, df = 2, p-value = 2.056e-10
```

Example: Nose bleeds and Alcohol

By direct calculation:

```
> #Form the fitted values from the row and column totals
> NB.rowtots<-apply(NoseBleeds,1,sum)
> NB.col tots<-apply(NoseBleeds,2,sum)
> NB.fitted<-outer(NB.rowtots,NB.col tots,"*")/sum(NoseBleeds)
> NB.fitted

+           [,1]      [,2]      [,3]
+ [1,] 54.23423 44.14414 41.62162
+ [2,] 31.76577 25.85586 24.37838

> (Xsq<-sum((NoseBleeds-NB.fitted)^2/NB.fitted))

+ [1] 44.61003

> (pval<-1-pchisq(Xsq,2))

+ [1] 2.056156e-10
```

Validity of the chi-squared test

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

$$np_i^{(0)} \quad i = 1, \dots, k$$

or

$$\hat{n}_{ij} \quad i = 1, \dots, r, j = 1, \dots, c$$

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

An Exact Test

If $r = c = 2$ we have a 2×2 table, and another **exact** test can be used: this 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**)
- R computes test statistic and p-value.

An Exact Test

The test statistic, T , is the (1,1) cell count used to test

H_0 : Factors are assigned independently.

If H_0 is **true**, then

$$T \sim \text{Hypergeometric}(n_{1.}, n_{2.}, n_{.1}, n_{.2})$$

This is a test carried out **conditional** on the row and column totals.

An Exact Test

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

An Exact Test

```
> JD<-matrix(c(1,8,5,2),2,2)
> chisq.test(JD, correct=FALSE)

+ Warning in chisq.test(JD, correct = FALSE): Chi-squared approximation may be incorrect

+
+ Pearson's Chi-squared test
+
+ data:  JD
+ X-squared = 6.1122, df = 1, p-value = 0.01343

> fisher.test(JD)

+
+ Fisher's Exact Test for Count Data
+
+ data:  JD
+ p-value = 0.03497
+ alternative hypothesis: true odds ratio is not equal to 1
+ 95 percent confidence interval:
+  0.0009525702 0.9912282442
+ sample estimates:
+ odds ratio
+ 0.06464255
```

An Exact Test

```
> fisher.test(JD, alternative='less')  
  
+  
+ Fisher's Exact Test for Count Data  
+  
+ data: JD  
+ p-value = 0.02448  
+ alternative hypothesis: true odds ratio is less than 1  
+ 95 percent confidence interval:  
+ 0.0000000 0.7290771  
+ sample estimates:  
+ odds ratio  
+ 0.06464255
```

An Exact Test

Here we have $T_{obs} = 1$. The possible values of T are

$$\{\max(0, n_{1.} - n_{2.}), \dots, \min(n_{1.}, n_{.1})\}$$

which here corresponds to the set

$$\{0, 1, 2, 3, 4, 5, 6\}$$

as

$$n_{.1} = 9 \quad n_{.2} = 7 \quad n_{1.} = 6 \quad n_{2.} = 10.$$

An Exact Test

In this Example

One-sided p-value : $P_0(T \leq 1)$

Two-sided p-value : $P_0(T \leq 1) + P_0(T \geq 6)$

That is

0.0244755 and 0.034965

respectively.

An Exact Test

Example: Juvenile Delinquency and Spectacle Wearing

Chi-squared Test:

$$\chi_{obs}^2 = 6.112$$

Compare with $\chi_{(r-1)(c-1)}^2 = \chi_1^2$; we have

$$\chi_{0.05,1}^2 = 3.841$$

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

Fisher's Exact Test: p-value is 0.035 (2-sided) or 0.024 (1-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 we collect and assess

- 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.

Case-Control Studies

Example: BCG Vaccination and Leprosy

In this example: we assess the possibility of an association between the two factors.

Disease Status : Leprosy Sufferer (Yes/No)

Factor : Vaccination Scar (Yes/No)

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

Odds Ratio

An alternative test statistic is based on the **odds ratio**

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

say. The test statistic is

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

where

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

Odds Ratio

That is,

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

If H_0

H_0 : No association between factor and disease status

is true, it follows that

$$Z \rightsquigarrow \text{Normal}(0, 1)$$

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

Odds Ratio

Example: BCG Vaccination and Leprosy

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

Therefore

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

and

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

Example: BCG Vaccination and Leprosy

Therefore

$$Z_{obs} = \frac{-0.671}{0.142} = -4.717$$

For a test at $\alpha = 0.05$, the two-sided critical values are ± 1.96 , so we **Reject** H_0 .

Odds Ratio

Example: Smoking and Lung Cancer

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

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

and

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

so

$$Z_{obs} = \frac{2.642}{0.735} = 3.590$$

Example: Smoking and Lung Cancer

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

Reject H_0

and report evidence for association.

Rank 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

Rank Tests

For example, if the data are

0.24 3.16 1.97 2.10 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.

Rank Tests

The statistical tasks are the same as in any other hypothesis testing setting, that is we need to

- define the **null hypothesis**
- define a suitable **test statistic**
- compute the **null distribution**
- define and utilize the **rejection region** to complete the test.

Single Population Tests

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

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

the **halfway** point of the distribution.

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

Single Population Tests

In a one-sample problem, we observe data

$$X_1, \dots, X_n$$

and we wish to carry out tests related to the underlying population.

If η is the population median, we wish to test, for example,

$$H_0 : \eta = \eta_0$$

for some fixed value η_0 .

Two Population Tests

In a two-sample problem, we observe data

$$X_{11}, \dots, X_{n1} \quad X_{12}, \dots, X_{n2}$$

from two populations.

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

One- and Two- sample tests

For the MWW test

- Label the samples so that $n_1 > n_2$ (i.e. sample 1 is the one with the larger sample size)
- The test is based on the **sum of the ranks** for the data from sample 2. We
 - ▶ Pool and sort the data, retaining the sample labels
 - ▶ Assign ranks in ascending order
 - ▶ The **rank sum** for sample 1 is R_1 , and the rank sum for sample 2 is denoted R_2 .

One- and Two- sample tests

Example: Mann-Whitney-Wilcoxon test statistic

SAMPLE 1	$n_1 = 4$	0.31	0.48	1.02	3.11
SAMPLE 2	$n_2 = 3$	0.16	0.20	1.97	

yields the following ranked data

SAMPLE	2	2	1	1	1	2	1
	0.16	0.20	0.31	0.48	1.02	1.97	3.11
RANK	1	2	3	4	5	6	7

Thus the rank sum for sample 1 and sample 2 are

$$R_1 = 3 + 4 + 5 + 7 = 19 \quad R_2 = 1 + 2 + 6 = 9.$$

One- and Two- sample tests

Two related test statistics can be used

- Wilcoxon Rank Sum Statistic

$$W = R_2$$

- Mann-Whitney U Statistic

$$U = R_2 - \frac{n_2(n_2 + 1)}{2}$$

One- and Two- sample tests

- **Large Sample Test** If $n_2 \geq 10$, a large sample test based on the Z statistic

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

can be used. Under $H_0 : \eta_1 = \eta_2$,

$$Z \sim \text{Normal}(0, 1)$$

and the test at $\alpha = 0.05$ uses the following critical values

$$H_a : \eta_1 > \eta_2 \quad \text{then} \quad C_R = -1.645$$

$$H_a : \eta_1 < \eta_2 \quad \text{then} \quad C_R = 1.645$$

$$H_a : \eta_1 \neq \eta_2 \quad \text{then} \quad C_R = \pm 1.960$$

One- and Two- sample tests

- **Small Sample Test** If $n_1 < 10$, an **exact** but more complicated test can be used.
 - ▶ The test statistic is R_2 .
 - ▶ The null distribution under the hypothesis $H_0 : \eta_1 = \eta_2$ can be computed, but it is complicated.

The table in McClave and Sincich gives the critical values (T_L and T_U) that determine the rejection region for different n_1 and n_2 values up to 10.

One- and Two- sample tests

- ▶ One-sided tests:

$$H_a : \eta_1 > \eta_2 \quad \text{Rejection Region is} \quad R_2 \leq T_L$$

$$H_a : \eta_1 < \eta_2 \quad \text{Rejection Region is} \quad R_2 \geq T_U$$

These are tests at the $\alpha = 0.025$ significance level.

- ▶ Two-sided tests:

$$H_a : \eta_1 \neq \eta_2$$

Rejection Region is

$$\{R_2 \leq T_L \text{ or } R_2 \geq T_U\}$$

This is a test at the $\alpha = 0.05$ significance level.

One- and Two- sample tests

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.

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

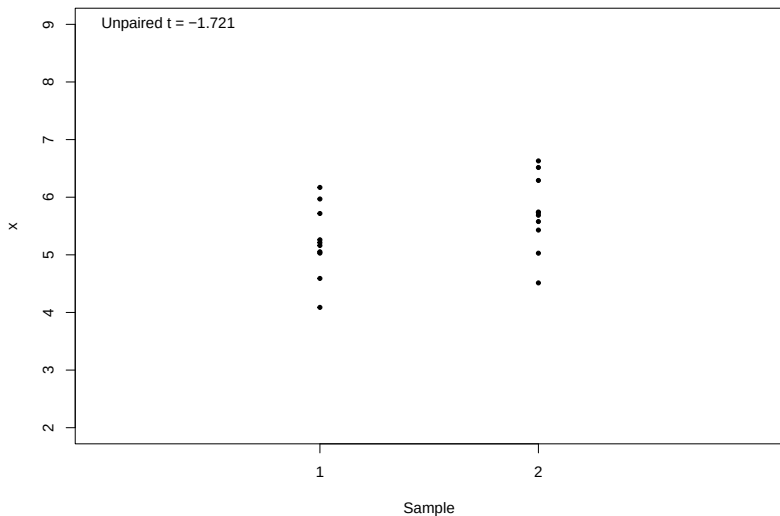
Wilcoxon Signed Rank Test

The Wilcoxon Signed Rank Test

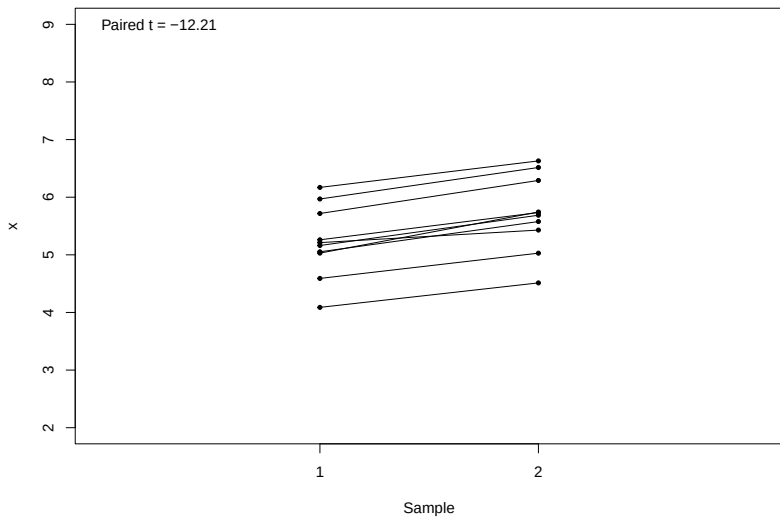
Remember:

- The **Wilcoxon** Test: a test for **independent** samples.
- The **Wilcoxon Signed Rank Test**: a test for **paired** data

Two sample test **without** Pairing



Two sample test **with** Pairing



The Wilcoxon Signed Rank Procedure

For paired data

- **pre**-treatment $(y_{11}, \dots, y_{n1})$
- **post**-treatment: $(y_{21}, \dots, y_{n2})$ data,

the procedure is as follows:

1. Form the **differences**

$$x_i = y_{i1} - y_{i2} \quad i = 1, \dots, n.$$

2. Form the **absolute values** $s_i = |x_i|$, for $i = 1, \dots, n$.
3. **Drop** every $x_i = 0$, leaving a **sample size of m** .

The Wilcoxon Signed Rank Procedure

4. Assign ranks $r_1, \dots, r_m$ to $s_1, \dots, s_m$ after sorting into ascending order.
5. Form

$$T_+ = \sum_{i=1}^m r_i z_i \qquad T_- = \sum_{i=1}^m r_i (1 - z_i)$$

where

$$z_i = \begin{cases} 1 & x_i > 0 \\ 0 & x_i < 0 \end{cases}$$

Computing the Null Distribution

The behaviour of T_+ and T_- can be predicted under the null hypothesis that the original data are generated from

continuous distributions

with the same median: we will never get $x_i = 0$.

- If K is the number of positive x_i values, then we have that $K \sim \text{Binomial}(n, 1/2)$, as in the sign test.
- Conditional on $K = k$, we can compute the distribution

$$P_0(T_+ = t | K = k)$$

by studying the behaviour of the ranks.

- We then use the Theorem of Total Probability to compute

$$P_0(T_+ = t)$$

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.

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.

Summary of the Non-Parametric Tests

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

Rank Correlation

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

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

where

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

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

Pearson Product Moment Coefficient of Correlation

Rank 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. Let the ranks for the data $x_1, \dots, x_n$ and $y_1, \dots, y_n$ be $u_1, \dots, u_n$ and $v_1, \dots, v_n$ 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**.

Rank 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. $u_1, \dots, u_n$ and $v_1, \dots, v_n$ can be preference ranks given by two observers.

Tests for r_S

To test

$$H_0 : \rho = 0$$

vs

$$(1) \quad H_a : \rho > 0$$

$$(2) \quad H_a : \rho < 0$$

$$(3) \quad H_a : \rho \neq 0$$

We may use r_S as a test statistic. The distribution of r_S under H_0 is tabulated in the textbook.

Tests for r_S

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

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

See textbook Table XI.

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 $x_1 \dots, x_{n_1}$ and $y_1 \dots, y_{n_2}$ (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.

Randomization and Permutation Tests

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

$$1 \quad 2 \quad 3 \quad \dots \quad n$$

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.

Randomization and Permutation Tests

3. If H_0

H_0 : No difference between popn. 1 and popn. 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.

Randomization and Permutation Tests

4. To obtain the exact distribution of W under H_0 , 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 $\{1, 2, 3, \dots, n\}$.

Randomization and Permutation Tests

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

$$\binom{6}{2} = \frac{6!}{2! 4!} = 15$$

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

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

Randomization and Permutation Tests

7. The rejection region and p-value are computed from the permutation null distribution. Let $W_i, i = 1, \dots, N$ 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

$$P_0(W \leq w) = \frac{\text{Number of } W_i \leq w}{N}$$

Randomization and Permutation Tests

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

$$P_0(W \in \mathcal{R}) = \frac{\text{Number of } W_i \in \mathcal{R}}{N} = \alpha$$

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**.

Randomization and Permutation Tests

Example: Simple Example

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

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

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

Randomization and Permutation Tests

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	

Randomization and Permutation Tests

There are 22 possible rank sums, $\{6, 7, 8, \dots, 25, 26, 27\}$; the number of times each sum is observed is displayed in the table, 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

Randomization and Permutation Tests

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

$$P_0(8 \leq W \leq 25) = 0.983 - 0.033 = 0.950$$

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: Placenta Permeability Data

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 $\{1, 2, 3, \dots, 15\}$ of size 5.

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

Placenta Permeability Data

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: Placenta Permeability Data

By inspection of the table, we see that

$$P_0(25 \leq W \leq 55) = 0.972 - 0.038 = 0.934$$

and

$$P_0(24 \leq W \leq 56) = 0.980 - 0.028 = 0.952$$

Placenta Permeability Data

Example: Placenta Permeability Data

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

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

and hence define the rejection region

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

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

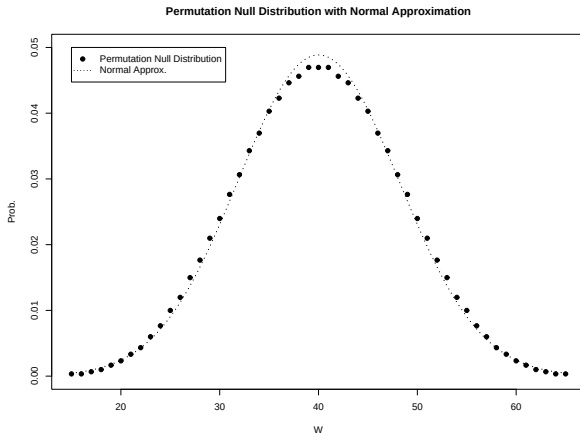
Thus in this example we **do not reject** the hypothesis of equal medians.

Normal Approximation

The Normal approximation to the exact null distribution is given by

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

Normal Approximation



Normal Approximation

The Normal approximation to the null distribution is quite good.

In fact, for most of the non-parametric tests, in large samples the Normal approximation yields a good approximation to the null distribution.

- Note that this Normal approximation is not the same as the Normality assumption we made in ANOVA and Regression settings
- The assumption for ANOVA and Regression related to each **data point**; the Normal approximation here relates to the behaviour of the **test statistic**.