# NON-PARAMETRIC STATISTICS
# ONE AND TWO SAMPLE TESTS

Non-parametric tests are normally based on **ranks** of the data samples, and test hypotheses relating to **quantiles** of the probability distribution representing the population from which the data are drawn. Specifically, tests concern the **population median**, $\eta$, where

$$\Pr[\text{ Observation } \le \eta\,] = \frac{1}{2}$$

The **sample median**, $x_{\text{MED}}$, is the mid-point of the sorted sample; if the data $x_1, \ldots, x_n$ are sorted into **ascending** order, then

$$x_{\text{MED}} = \begin{cases} x_m & n \text{ odd}, n = 2m+1 \\ \dfrac{x_m + x_{m+1}}{2} & n \text{ even}, n = 2m \end{cases}$$

## 1 ONE SAMPLE TEST FOR MEDIAN: THE SIGN TEST

For a single sample of size $n$, to test the hypothesis $\eta = \eta_0$ for some specified value $\eta_0$ we use the **Sign Test.**. The test statistic $S$ depends on the alternative hypothesis, $H_a$.

(a) For **one-sided** tests, to test

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta > \eta_0 \end{aligned}$$

we define test statistic $S$ by

$$S = \text{Number of observations } \textbf{greater than } \eta_0$$

whereas to test

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta < \eta_0 \end{aligned}$$

we define $S$ by

$$S = \text{Number of observations } \textbf{less than } \eta_0$$

If $H_0$ is **true**, it follows that

$$S \sim \text{Binomial}\left(n, \frac{1}{2}\right)$$

The $p$-value is defined by

$$p = \Pr[X \ge S]$$

where $X \sim \text{Binomial}(n, 1/2)$. The rejection region for significance level $\alpha$ is defined implicitly by the rule

$$\text{Reject } H_0 \text{ if } \alpha \ge p.$$

The Binomial distribution is tabulated on pp 885-888 of McClave and Sincich.

(b) For a **two-sided** test,

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_a &: \eta \neq \eta_0 \end{aligned}$$

we define the test statistic by

$$S = \max\{S_1, S_2\}$$

where $S_1$ and $S_2$ are the counts of the number of observations less than, and greater than, $\eta_0$ respectively. The $p$-value is defined by

$$p = 2\Pr[X \geq S]$$

where $X \sim \text{Binomial}(n, 1/2)$.

**Notes :**

1. The only assumption behind the test is that the data are drawn independently from a continuous distribution.

2. If any data are equal to $\eta_0$, we **discard** them before carrying out the test.

3. **Large sample approximation.** If $n$ is large (say $n \geq 30$), and $X \sim \text{Binomial}(n, 1/2)$, then it can be shown that

   $$X \stackrel{\cdot}{\sim} \text{Normal}(np, np(1-p))$$

   Thus for the sign test, where $p = 1/2$, we can use the test statistic

   $$Z = \frac{S - \dfrac{n}{2}}{\sqrt{n \times \dfrac{1}{2} \times \dfrac{1}{2}}} = \frac{S - \dfrac{n}{2}}{\sqrt{n} \times \dfrac{1}{2}}$$

   and note that if $H_0$ is true,

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

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

   $$\begin{array}{lll} H_a : \eta > \eta_0 & \text{then} & C_R = 1.645 \\ H_a : \eta < \eta_0 & \text{then} & C_R = -1.645 \\ H_a : \eta \neq \eta_0 & \text{then} & C_R = \pm 1.960 \end{array}$$

4. For the large sample approximation, it is common to make a **continuity correction**, where we replace $S$ by $S - 1/2$ in the definition of $Z$

   $$Z = \frac{\left(S - \dfrac{1}{2}\right) - \dfrac{n}{2}}{\sqrt{n} \times \dfrac{1}{2}}$$

Tables of the standard Normal distribution are given on p 894 of McClave and Sincich.

## 2 TWO SAMPLE TESTS FOR INDEPENDENT SAMPLES: THE MANN-WHITNEY-WILCOXON TEST

For a two **independent** samples of size $n_1$ and $n_2$, to test the hypothesis of **equal population medians**

$$\eta_1 = \eta_2$$

we use the **Wilcoxon Rank Sum Test**, or an equivalent test, the **Mann-Whitney U Test**; we refer to this as the

**Mann-Whitney-Wilcoxon (MWW) Test**

By convention it is usual to formulate the test statistic in terms of the **smaller** sample size. Without loss of generality, we label the samples such that

$$n_1 > n_2.$$

The test is based on the **sum of the ranks** for the data from sample 2.

**EXAMPLE :** $n_1 = 4, n_2 = 3$ yields the following ranked data

| | | | | |
|---|---|---|---|---|
| SAMPLE 1 | 0.31 | 0.48 | 1.02 | 3.11 |
| SAMPLE 2 | 0.16 | 0.20 | 1.97 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 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 is

$$R_1 = 3 + 4 + 5 + 7 = 19$$

and the rank sum for sample 2 is

$$R_2 = 1 + 2 + 6 = 9.$$

Let $\eta_1$ and $\eta_2$ denote the medians from the two distributions from which the samples are drawn. We wish to test

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

Two related test statistics can be used

- **Wilcoxon Rank Sum Statistic**

$$W = R_2$$

- **Mann-Whitney U Statistic**

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

We again consider three alternative hypotheses:

$$\begin{aligned} H_a &: \eta_1 < \eta_2 \\ H_a &: \eta_1 > \eta_2 \\ H_a &: \eta_1 = \eta_2 \end{aligned}$$

and define the rejection region separately in each case.

**Large Sample Test**

If $n_2 \geq 10$, a large sample test based on the $Z$ statistic

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

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

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

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

$$\begin{array}{lll} H_a : \eta_1 > \eta_2 & \text{then} & C_R = -1.645 \\ H_a : \eta_1 < \eta_2 & \text{then} & C_R = 1.645 \\ H_a : \eta_1 \neq \eta_2 & \text{then} & C_R = \pm 1.960 \end{array}$$

**Small Sample Test**

If $n_1 < 10$, an **exact** but more complicated test can be used. The test statistic is $R_2$ (the sum of the ranks for sample 2). The null distribution under the hypothesis $H_0 : \eta_1 = \eta_2$ can be computed, but it is complicated.

The table on p. 832 of 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-sided tests:**

$$\begin{array}{lll} H_a : \eta_1 > \eta_2 & \text{Rejection Region is} & R_2 \leq T_L \\ H_a : \eta_1 < \eta_2 & \text{Rejection Region is} & R_2 \geq T_U \end{array}$$

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

- **Two-sided tests:**

$$H_a : \eta_1 \neq \eta_2 \quad \text{Rejection Region is} \quad R_2 \leq T_L \text{ or } R_2 \geq T_U$$

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

**Notes :**

1. The only assumption is are needed for the test to be valid is that the samples are independently drawn from two continuous distributions.

2. The sum of the ranks across **both** samples is

$$R_1 + R_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2}$$

3. If there are **ties** (equal values) in the data, then the rank values are replaced by **average** rank values.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DATA VALUE | 0.16 | 0.20 | 0.31 | 0.31 | 0.48 | 1.97 | 3.11 |
| ACTUAL RANK | 1 | 2 | 3 | 3 | 5 | 6 | 7 |
| AVERAGE RANK | 1 | 2 | 3.5 | 3.5 | 5 | 6 | 7 |

## EXAMPLES

**EXAMPLE 1: Sign Test: Water Content Example**

The following data are measurements of percentage water content of soil samples collected by two experimenters. We wish to test the hypothesis

$$H_0 \; : \; \eta = 9.0$$

for each experiment.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Experimenter 1: | $n = 10$ | 5.5 | 6.0 | 6.5 | 7.6 | 7.6 | 7.7 | 8.0 | 8.2 | 9.1 | 15.1 |
| Experimenter 2: | $n = 20$ | 5.6 | 6.1 | 6.3 | 6.3 | 6.5 | 6.6 | 7.0 | 7.5 | 7.9 | 8.0 |
| | | 8.0 | 8.1 | 8.1 | 8.2 | 8.4 | 8.5 | 8.7 | 9.4 | 14.3 | 26.0 |

To perform the test, we need tables of the Binomial distribution with $p = 1/2$. The individual probabilities are given by the formula

$$\Pr[X = x] = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \frac{1}{2^n} = \frac{n!}{x!(n-x)!} \frac{1}{2^n} \qquad x = 0, 1, \ldots, n$$

We test at the $\alpha = 0.05$ level. For the first experiment, with $n = 10$:

- For a test against the alternative hypothesis

  $$H_a \; : \; \eta > 9.0$$

  the test statistic is

  $$S = \text{Number of observations } \textbf{greater than } 9 \qquad \therefore \qquad S = 2$$

  and the $p$-value is

  $$p = \Pr[X \geq 2] = 1 - \Pr[X < 2] = 1 - \Pr[X = 0] - \Pr[X = 1] = 0.9893$$

  so we **do not** reject $H_0$ in favour of this $H_a$.

- For a test against the alternative hypothesis

  $$H_a \; : \; \eta < 9.0$$

  the test statistic is

  $$S = \text{Number of observations } \textbf{less than } 9 \qquad \therefore \qquad S = 8$$

  and the $p$-value is

  $$p = \Pr[X \geq 8] = \Pr[X = 8] + \Pr[X = 9] + \Pr[X = 10] = 0.0547$$

  so we **do not** reject $H_0$ in favour of this $H_a$.

- For a test against the alternative hypothesis

  $$H_a \; : \; \eta \neq 9.0$$

  the test statistic is

  $$S = \max\{S_1, S_2\} = \max\{2, 8\} = 8$$

  and the $p$-value is

  $$p = 2\Pr[X \geq 8] = 2(\Pr[X = 8] + \Pr[X = 9] + \Pr[X = 10]) = 0.1094$$

  so we **do not** reject $H_0$ in favour of this $H_a$.

For the second experiment, with $n = 20$:

- For a test against the alternative hypothesis $H_a : \eta > 9.0$, the test statistic is $S = 3$. The $p$-value is therefore

$$p = \Pr[X \geq 3] = 1 - \Pr[X < 3] = 1 - \Pr[X = 0] - \Pr[X = 1] - \Pr[X = 2] = 0.9998.$$

  so we **do not** reject $H_0$ in favour of this $H_a$.

- For a test against the alternative hypothesis $H_a : \eta < 9.0$, the test statistic $S = 17$. The $p$-value is therefore

$$p = \Pr[X \geq 17] = \Pr[X = 17] + \Pr[X = 18] + \Pr[X = 19] + \Pr[X = 20] = 0.0013.$$

  so we **do** reject $H_0$ in favour of this $H_a$.

- For a test against the alternative hypothesis $H_a : \eta \neq 9.0$, the test statistic is $S = \max\{S_1, S_2\} = \max\{3, 17\} = 17$. The $p$-value is therefore

$$p = 2\Pr[X \geq 17] = 2(\Pr[X = 17] + \Pr[X = 18] + \Pr[X = 19] + \Pr[X = 20]) = 0.0026.$$

  so we **do** reject $H_0$ in favour of this $H_a$.

This test can be implemented using SPSS, using the

*Analyze* $\rightarrow$ *Nonparametric Tests* $\rightarrow$ *Binomial*

pulldown menus. The test can be carried out by

(a) Selecting the *test variable* from the variables list
(b) Set the *Cut Point* equal to $\eta_0 = 9$.

A **two-sided** test is carried out at the $\alpha = 0.05$ level. The SPSS output is presented below for the two experiments in turn:

**Binomial Test**

| | | Category | N | Observed Prop. | Test Prop. | Exact Sig. (2-tailed) |
|---|---|---|---|---|---|---|
| % Water content | Group 1 | <= 9 | 8 | .80 | .50 | .109 |
| | Group 2 | > 9 | 2 | .20 | | |
| | Total | | 10 | 1.00 | | |

**Binomial Test**

| | | Category | N | Observed Prop. | Test Prop. | Exact Sig. (2-tailed) |
|---|---|---|---|---|---|---|
| % Water content | Group 1 | <= 9 | 17 | .85 | .50 | .003 |
| | Group 2 | > 9 | 3 | .15 | | |
| | Total | | 20 | 1.00 | | |

**EXAMPLE 2: Mann-Whitney-Wilcoxon Test: Low Birthweight Example**
The birthweights (in grammes) of babies born to two groups of mothers A and B are displayed below: Thus $n_1 = 9, n_2 = 8$. From this sample (which has ties, so we need to use average ranks), we find that

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Group A : | $n = 9$ | 2164 | 2600 | 2184 | 2080 | 1820 | 2496 | 2184 | 2080 | 2184 |
| Group B : | $n = 8$ | 2576 | 3224 | 2704 | 2912 | 2444 | 3120 | 2912 | 3848 | |

$$R_1 = 48 \qquad R_2 = 105$$

so that the two statistics are

$$\text{Wilcoxon} \quad W \quad = R_2 = 105$$

$$\text{Mann-Whitney} \quad U \quad = R_2 - \frac{n_2(n_2+1)}{2} = 105 - 36 = 69$$

- For the **small sample** test, from tables on p832 in McClave and Sincich, we find

  $$T_L = 51 \qquad T_U = 93$$

  Thus $W > 93$, so we

  **Do not** reject $H_0$ against $H_a \ : \ \eta_1 > \eta_2$ as $W = R_2 > T_L$
  **Reject** $H_0$ against $H_a \ : \ \eta_1 < \eta_2$ as $W = R_2 > T_U$
  **Reject** $H_0$ against $H_a \ : \ \eta_1 \neq \eta_2$ as $W = R_2 > T_U$

  Note that the one-sided tests are carried out at $\alpha = 0.025$, the two sided test is carried out at $\alpha = 0.05$.

- For the **large sample** test, we find

  $$Z = \frac{U - \dfrac{n_1 n_2}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = 3.175$$

  Thus we

  **Do not** reject $H_0$ against $H_a \ : \ \eta_1 > \eta_2$ as $Z > C_R = -1.645$
  **Reject** $H_0$ against $H_a \ : \ \eta_1 < \eta_2$ as $Z > C_R = 1.645$
  **Reject** $H_0$ against $H_a \ : \ \eta_1 \neq \eta_2$ as $Z > C_{R_2} = 1.960$

  All tests are carried out at $\alpha = 0.05$.

This test can be implemented using SPSS, using the

*Analyze* → *Nonparametric Tests* → *Two Independent Samples*

pulldown menus. Note, however, that SPSS uses different rules for defining the test statistics, although it yields the same conclusions for a two-sided test.

**EXAMPLE 3: Mann-Whitney-Wilcoxon Test: Treadmill Test Example**
The treadmill stress test times (in seconds) of two groups of patients (disease group and healthy controls) are displayed below:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Disease : | $n = 10$ | 864 | 636 | 638 | 708 | 786 | 600 | 1320 | 750 | 594 | 750 |
| Healthy : | $n = 8$ | 1014 | 684 | 810 | 990 | 840 | 978 | 1002 | 1110 | | |

Thus $n_1 = 10, n_2 = 8$. From this sample (which has ties, so we need to use average ranks), we find that

$$R_1 = 70 \qquad R_2 = 101$$

so that the two statistics are

$$\text{Wilcoxon} \quad W \quad = R_2 = 101$$

$$\text{Mann-Whitney} \quad U \quad = R_2 - \frac{n_2(n_2+1)}{2} = 101 - 36 = 65$$

- For the **small sample** test, from tables on p832 in McClave and Sincich, we find

$$T_L = 54 \qquad T_U = 98$$

  Thus $W > 98$, so we

  **Do not** reject $H_0$ against $H_a \; : \; \eta_1 > \eta_2$ as $W = R_2 > T_L$
  **Reject** $H_0$ against $H_a \; : \; \eta_1 < \eta_2$ as $W = R_2 > T_U$
  **Reject** $H_0$ against $H_a \; : \; \eta_1 \neq \eta_2$ as $W = R_2 > T_U$

  Again, the one-sided tests are carried out at $\alpha = 0.025$, the two sided test is carried out at $\alpha = 0.05$.

- For the **large sample** test, we find

$$Z = \frac{U - \dfrac{n_1 n_2}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = 2.221$$

  Thus we

  **Do not** reject $H_0$ against $H_a \; : \; \eta_1 > \eta_2$ as $Z > C_R = -1.645$
  **Reject** $H_0$ against $H_a \; : \; \eta_1 < \eta_2$ as $Z > C_R = 1.645$
  **Reject** $H_0$ against $H_a \; : \; \eta_1 \neq \eta_2$ as $Z > C_{R_2} = 1.960$

  All tests are carried out at $\alpha = 0.05$.

## 3 TWO DEPENDENT SAMPLES: WILCOXON SIGNED RANK TEST

Data collected from the same experimental units are in general **dependent**. For example, if data are collected on two occasions (time 1 and time 2, or before and after treatment) from the same $n$ individuals, then the resulting data samples $(y_{11}, \ldots, y_{n1})$ and $(y_{12}, \ldots, y_{n2})$ are dependent. Such data are often referred to as **paired**. We wish to test whether there is a significant change across the two measurements.

For a **parametric** test, we typically assume that the within-individual differences

$$x_i = y_{i1} - y_{i2} \qquad i = 1, \ldots, n$$

are **Normally** distributed, and test the hypothesis that the mean difference $\mu$ is zero

$$H_0 \, : \, \mu = 0$$

using a one-sample $Z$-test ($\sigma$ known) or $T$-test ($\sigma$ unknown), with statistic

$$z = \frac{\overline{x}}{\sigma/\sqrt{n}} \qquad \text{or} \qquad t = \frac{\overline{x}}{s/\sqrt{n}}$$

distributed as Normal$(0, 1)$ or Student$(n-1)$ respectively.

For a **non-parametric** test, we can use the **Wilcoxon Signed Rank** test, which proceeds as follows:

1. Compute the within-individual differences

$$x_i = y_{i1} - y_{i2} \qquad i = 1, \ldots, n$$

   If any $x_i = 0$, then that data point is discarded and the sample size adjusted.

2. Sort the **absolute values** $s_1, \ldots, s_n$ of $x_1, x_2, \ldots, x_n$ into **ascending** order, and assign ranks 1 up to $n$. If there are ties, assign **average** ranks.

3. Form the two rank sums $T_+$ and $T_-$, where

$$\begin{aligned} T_+ &= \text{Sum of ranks for those } x_i > 0 \\ T_- &= \text{Sum of ranks for those } x_i < 0 \end{aligned}$$

The test statistic is a function of these rank sums. Heuristically, if the statistic $T_+$ is large and $T_-$ is small, this implies that the experimental units where $y_{i1} > y_{i2}$ have a **larger** (in magnitude) difference than those where $y_{i1} < y_{i2}$. This indicates an overall **decrease** between the first and second measurements. Conversely, if the statistic $T_-$ is large and $T_+$ is small, this implies that the experimental units where $y_{i2} > y_{i1}$ have a **larger** (in magnitude) difference than those where $y_{i2} < y_{i1}$. This indicates an overall **increase** between the first and second measurements.

We test the null hypothesis

$$H_0 \, : \, \text{No change between first and second measurements}$$

against the three alternative hypotheses

$$\begin{aligned} (1)\ H_a &: \text{Significant } \textbf{decrease} \text{ between first and second measurements} \\ (2)\ H_a &: \text{Significant } \textbf{increase} \text{ between first and second measurements} \\ (3)\ H_a &: \text{Significant } \textbf{change} \text{ between first and second measurements} \end{aligned}$$

To test $H_0$ vs (1), we perform a one-sided test using the statistic $T_-$; the critical value in the test is denoted $T_0$, and is determined by the table on p. 839 of McClave and Sincich:

If $T_- \leq T_0$, we **reject** $H_0$ in favour of $H_a$ (1)

To test $H_0$ vs (2), we perform a one-sided test using the statistic $T_+$; the critical value is $T_0$ and

If $T_+ \leq T_0$, we **reject** $H_0$ in favour of $H_a$ (2)

To test $H_0$ vs (3), we perform a two-sided test using the statistic $T = \min\{T_-, T_+\}$; the critical value is $T_0$ and

If $T \leq T_0$, we **reject** $H_0$ in favour of $H_a$ (3)

**Notes :**

1. The only assumption behind the test is that the difference data $x_i$ are drawn independently from a continuous distribution.

2. **Large Sample Test:** For $n \geq 25$, we can use a large sample version of the test based on $T_+$, and the $Z$ statistic

$$Z = \frac{T_+ - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

If $H_0$ is **true**, then $Z \stackrel{\cdot}{\sim}$ Normal$(0, 1)$, so that the test at $\alpha = 0.05$ uses the following critical values

| | | |
|---|---|---|
| For $H_a$ (1) | use | $C_R = 1.645$ |
| For $H_a$ (2) | use | $C_R = -1.645$ |
| For $H_a$ (3) | use | $C_R = \pm 1.960$ |

**EXAMPLE 1: Haemodialysis Data**
The following data are measurements of the heparin cofactor II (HCII) to plasma protein ratios in a group of patients at baseline and five months after haemodialysis.

Reference: Toulon, P *et al.* (1987) Antithrombin III and heparin cofactor II in patients with chronic renal failure undergoing regular hemodialysis, *Thrombosis and Haemostasis*, **3**;57(3): pp263-8.

| Patient | Before | After | | | | | |
|---|---|---|---|---|---|---|---|
| | $y_{i1}$ | $y_{i2}$ | $x_i$ | $s_i$ | Rank | Ave. Rank | |
| 1 | 2.11 | 2.15 | -0.04 | 0.04 | 3 | 3.5 | |
| 2 | 1.85 | 2.11 | -0.26 | 0.26 | 10 | 10.0 | |
| 3 | 1.82 | 1.93 | -0.11 | 0.11 | 8 | 8.0 | |
| 4 | 1.75 | 1.83 | -0.08 | 0.08 | 6 | 6.0 | |
| 5 | 1.54 | 1.90 | -0.36 | 0.36 | 11 | 11.0 | |
| 6 | 1.52 | 1.56 | -0.04 | 0.04 | 3 | 3.5 | |
| 7 | 1.49 | 1.44 | 0.05 | 0.05 | 5 | 5.0 | |
| 8 | 1.44 | 1.43 | 0.01 | 0.01 | 1 | 1.5 | |
| 9 | 1.38 | 1.28 | 0.10 | 0.10 | 7 | 7.0 | |
| **10** | **1.30** | **1.30** | **0.00** | **0.00** | **-** | **-** | **OMIT** |
| 11 | 1.20 | 1.21 | -0.01 | 0.01 | 1 | 1.5 | |
| 12 | 1.19 | 1.30 | -0.11 | 0.11 | 9 | 9.0 | |
| | | | | | | $T_+ = 13.5$ | |
| | | | | | | $T_- = 52.5$ | |

From the table on p 839, for $n = 12 - 1 = 11$, we find that the $\alpha = 0.025/0.05$ (one/two-sided) significance level critical value is $T_0 = 11$. Thus using $T_+$, we **cannot reject** either of the null hypotheses (2) and (3), as $T_+ > T_0$ . Note that $Z = -1.734$, so if the approximation was valid, we would be able to reject (2) at $\alpha = 0.05$.

# 4 THREE OR MORE INDEPENDENT SAMPLES: THE KRUSKAL-WALLIS AND FRIEDMAN TESTS

We now seek non-parametric tests that can be used for multiple independent samples, such as those found in the Completely Randomized Design (CRD) and Randomized Block Design (RBD) described in the ANOVA section. The non-parametric equivalents of the Fisher-F tests for these two designs are

- The **Kruskal-Wallis *H* test** for a Completely Randomized Design
- **Friedman's test** for a Randomized Block Design

## 4.1 Kruskal-Wallis Test

In a CRD, we have $k$ independent groups, corresponding to $k$ different treatments, with sample sizes $n_1, \ldots, n_k$. Let $n = n_1 + \cdots + n_k$. To compute the test statistic, $H$, we

1. Pool the data, sort them into ascending order, and assign ranks. If there are ties in the data, then average ranks are used.
2. For $j = 1, \ldots, k$, compute the rank sum $R_j$

$$R_j = \text{Sum of ranks for data from sample } j.$$

To test the hypothesis

$$\begin{array}{lcl} H_0 & : & \text{No difference between the population distributions of the } k \text{ groups} \\ H_a & : & \text{At least two population distributions different} \end{array}$$

the test statistic is

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

If $H_0$ is **true**, then for large $n$,

$$H \stackrel{\sim}{\sim} \text{Chisquared}(k-1).$$

**Notes :**

1. The test assumes that the $k$ samples are independently drawn from continuous populations.
2. For the approximation to be valid, there should be at least **five** observations in each sample, and the number of ties should be small.

**EXAMPLE 2: Mucociliary efficiency data**
The data are measures of mucociliary efficiency from the rate of removal of dust in normal subjects (Group 1), subjects with obstructive airway disease (Group 2), and subjects with asbestosis (Group 3).

Reference: Myles Hollander, M and Douglas A. Wolfe (1973), *Nonparametric statistical inference*, New York: John Wiley & Sons. pp115-120.

| Group | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 2.9 | 3.0 | 2.5 | 2.6 | 3.2 | 3.8 | 2.7 | 4.0 | 2.4 | 2.8 | 3.4 | 3.7 | 2.2 | 2.0 |
| Rank | 8 | 9 | 4 | 5 | 10 | 13 | 6 | 14 | 3 | 7 | 11 | 12 | 2 | 1 |

Hence $R_1 = 36$, $R_2 = 36$ and $R_3 = 33$, and the test statistic $H = 0.7714$. To complete the test, we compare with the $\alpha = 0.05$ quantile of the $\text{Chisquared}(k-1) = \text{Chisquared}(2)$ distribution. We have

$$\text{Chisq}_{0.05}(2) = 5.99 > H \qquad \therefore \qquad \textbf{No evidence to reject } H_0$$

and a $p$-value of $p = 0.680$.

## 4.2 Friedman Test

In a RBD, we have $k$ treatment groups, and a blocking factor. For example, we might have $k$ repeated measurements on the same $b$ experimental units, and $n = bk$ observations in total. To compute the test statistic, $F_r$, we proceed as follows.

1. **Within each block separately**, sort the $k$ data values into ascending order, and assign ranks. If there are ties in the data, then average ranks are used.

2. For $j = 1, \ldots, k$, compute the rank sum $R_j$

$$R_j = \text{Sum of ranks for data from } \textbf{treatment } j.$$

To test the hypothesis

$$\begin{aligned} H_0 &: \text{No difference between the population distributions of the } k \text{ treatment groups} \\ H_a &: \text{At least two population distributions different} \end{aligned}$$

the test statistic is

$$F_r = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_j^2 - 3b(k+1)$$

If $H_0$ is **true**, then for large $n$,

$$F_r \stackrel{\cdot}{\sim} \text{Chisq}(k-1)$$

**Notes :**

1. The test assumes that the data are drawn independently from continuous populations, with random assignment of treatments within blocks.

2. For the approximation to be valid, it is recommended that $b$ or $k$ is at least five, and the number of ties should be small.

**EXAMPLE 3: Skin potential under hypnosis**
A study was conducted to investigate whether hypnosis has the same effect on skin potential for four different emotions. Eight subjects were asked to display fear, joy, sadness and calmness under hypnosis, and the resulting skin potential (measured in millivolts) was recorded for each emotion. Thus in this experiment, $b = 8$ and $k = 4$.

| | Fear | | Joy | | Sadness | | Calmness | |
|---|---|---|---|---|---|---|---|---|
| Subject | $y$ | Rank | $y$ | Rank | $y$ | Rank | $y$ | Rank |
| 1 | 23.1 | 4 | 22.7 | 3 | 22.5 | 1 | 22.6 | 2 |
| 2 | 57.6 | 4 | 53.2 | 2 | 53.7 | 3 | 53.1 | 1 |
| 3 | 10.5 | 3 | 9.7 | 2 | 10.8 | 4 | 8.3 | 1 |
| 4 | 23.6 | 4 | 19.6 | 3 | 21.1 | 2 | 21.6 | 1 |
| 5 | 11.9 | 1 | 13.8 | 4 | 13.7 | 3 | 13.3 | 2 |
| 6 | 54.6 | 4 | 47.1 | 3 | 39.2 | 2 | 37.0 | 1 |
| 7 | 21.0 | 4 | 13.6 | 1 | 13.7 | 2 | 14.8 | 3 |
| 8 | 20.3 | 3 | 23.6 | 4 | 16.3 | 2 | 14.8 | 1 |
| Rank Sum | | 27 | | 20 | | 19 | | 14 |

Thus the within-treatment rank sums are $R_1 = 27, R_2 = 20, R_3 = 19$ and $R_4 = 14$ and thus $F_r = 6.45$. To complete the test, we compare with the $\alpha = 0.05$ quantile of the

$$\text{Chisquared}(k-1) = \text{Chisquared}(3)$$

distribution. We have

$$\text{Chisq}_{0.05}(3) = 7.81 > F_r \qquad \therefore \qquad \textbf{No evidence to reject } H_0$$

and a $p$-value of $p = 0.092$.

## 5 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.

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
$$H_0 \ : \ \text{No difference between population 1 and population2}$$
were **true**, then there should be **no pattern** in the group labels when sorted into ascending order; 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 $H_0$ (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 $n = 6$ 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** $\alpha$ as the distribution of $W$ is **discrete**.

**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 $\{1, 2, 3, \ldots, 10\}$ 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:

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

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

## 6 SPEARMAN'S RANK CORRELATION

A measure of association for two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Pearson Product Moment Correlation Coefficient**, $r$, where

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

where

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

This quantity measures the **linear** association between the $X$ and $Y$ variables.

A measure of the potentially non-linear association between the samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Spearman Rank Correlation Coefficient**, $r_S$, which computes the correlation between the **ranks** of the data.

The Spearman Rank Correlation Coefficient is computed as follows:

1. Assign ranks $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ to the data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ separately by sorting each sample into ascending order and assigning the ranks in order.
2. Compute $r_S$ as

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

   where

   $$SS_{uu} = \sum_{i=1}^{n}(u_i - \overline{u})^2 \qquad SS_{vv} = \sum_{i=1}^{n}(v_i - \overline{v})^2 \qquad SS_{uv} = \sum_{i=1}^{n}(u_i - \overline{u})(v_i - \overline{v})$$

If there are no ties in the data, then

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

where

$$d_i = u_i - v_i \qquad i = 1, \ldots, n$$

**Tests for $r_S$ :** If the population correlation is $\rho$, then we may test the hypothesis

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

against the hypotheses

$$\begin{aligned} (1)\; H_a &\;:\; \rho > 0 \\ (2)\; H_a &\;:\; \rho < 0 \\ (3)\; H_a &\;:\; \rho \neq 0 \end{aligned}$$

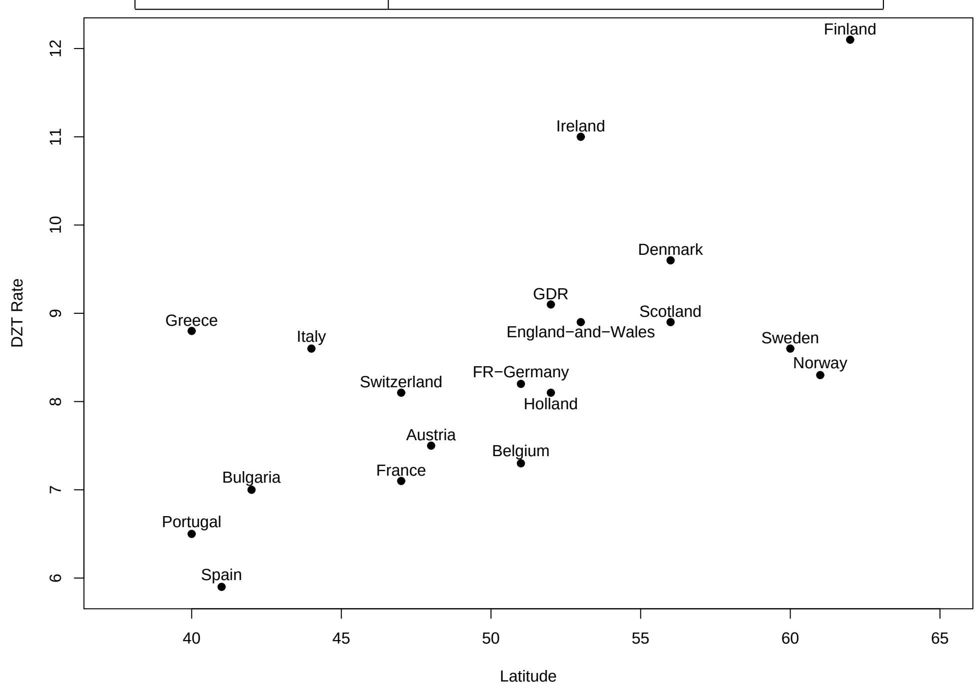
using the table of the null distribution on p 894 of McClave and Sincich. If $\text{Spearman}_\alpha$ is the $\alpha$ tail quantile of the null distribution, we have the following rejection regions:

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

**EXAMPLE : Latitude and dizygotic twinning rates**
The relationship between the geographical latitude of a country and its dizygotic twinning (DZT) rate is to be investigated. The data are presented and plotted below.

Reference: James, W.H. (1985) Dizygotic twinning, birth weight and latitude, *Annals of Human Biology*, **12**, 5, pp. 441-447.

| Country | Latitude $x$ | Rank $u$ | DZT Rate $y$ | Rank $v$ |
|---|---|---|---|---|
| Portugal | 40 | 1.5 | 6.5 | 2.0 |
| Greece | 40 | 1.5 | 8.8 | 13.0 |
| Spain | 41 | 3.0 | 5.9 | 1.0 |
| Bulgaria | 42 | 4.0 | 7.0 | 3.0 |
| Italy | 44 | 5.0 | 8.6 | 11.5 |
| France | 47 | 6.5 | 7.1 | 4.0 |
| Switzerland | 47 | 6.5 | 8.1 | 7.5 |
| Austria | 48 | 8.0 | 7.5 | 6.0 |
| Belgium | 51 | 9.5 | 7.3 | 5.0 |
| FR Germany | 51 | 9.5 | 8.2 | 9.0 |
| Holland | 52 | 11.5 | 8.1 | 7.5 |
| GDR | 52 | 11.5 | 9.1 | 16.0 |
| England & Wales | 53 | 13.5 | 8.9 | 14.5 |
| Ireland | 53 | 13.5 | 11.0 | 18.0 |
| Scotland | 56 | 15.5 | 8.9 | 14.5 |
| Denmark | 56 | 15.5 | 9.6 | 17.0 |
| Sweden | 60 | 17.0 | 8.6 | 11.5 |
| Norway | 61 | 18.0 | 8.3 | 10.0 |
| Finland | 62 | 19.0 | 12.1 | 19.0 |




For these data

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}\,SS_{vv}}} = \frac{384.5}{\sqrt{567 \times 568.5}} = 0.677 \qquad r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}} = \frac{118.4}{\sqrt{866.105 \times 38.88}} = 0.645$$

indicating a strong positive association.