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

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

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**

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

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 TEST** or the **MANN-WHITNEY U TEST**.

SEE HANDOUT

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**

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

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 TEST** or the **MANN-WHITNEY U TEST**.

SEE HANDOUT

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**, η, where

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

The **sample median**, x_{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 η_0 we use the **Sign Test.**. The test statistic S depends on the alternative hypothesis, H_a.

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

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

we define $S =$ Number of observations **greater than** η_0, whereas to test

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

we define $S =$ Number of observations **less than** η_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

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

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

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

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, η_0 respectively. The p-value is defined by

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

where $X \sim$ 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 η_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

$$X \mathrel{\dot\sim} \text{Normal}(np, np(1-p))$$

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

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

and note that if H_0 is true,

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

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

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

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 : $n_1 = 4, n_2 = 3$

SAMPLE 1	0.31	0.48	1.02	3.11
SAMPLE 2	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 is

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

and the rank sum for sample 2 is

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

Let η_1 and η_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:

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

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

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

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

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

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:**

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

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

- **Two-sided tests:**

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

  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

NON-PARAMETRIC STATISTICS: ONE AND TWO SAMPLE TESTS
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 → *Nonparametric Tests* → *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

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	

sample (which has ties, so we need to use average ranks), we find that

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

so that the two statistics are

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

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

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

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

Correction

Thus $W > 93$, so we

$$\begin{array}{rl} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a : \eta_1 > \eta_2 & \text{as} \quad W = R_2 > T_L \\ \textbf{Reject } H_0 \text{ against } H_a : \eta_1 < \eta_2 & \text{as} \quad W = R_2 > T_U \\ \textbf{Reject } H_0 \text{ against } H_a : \eta_1 \neq \eta_2 & \text{as} \quad W = R_2 > T_U \end{array}$$

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

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

Correction

Thus we

$$\begin{array}{rl} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a : \eta_1 > \eta_2 & \text{as} \quad Z > C_R = -1.645 \\ \textbf{Reject } H_0 \text{ against } H_a : \eta_1 < \eta_2 & \text{as} \quad Z > C_R = 1.645 \\ \textbf{Reject } H_0 \text{ against } H_a : \eta_1 \neq \eta_2 & \text{as} \quad Z > C_{R_2} = 1.960 \end{array}$$

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:

$$\begin{array}{llllllllllll} \text{Disease}: & n=10 & 864 & 636 & 638 & 708 & 786 & 600 & 1320 & 750 & 594 & 750 \\ \text{Healthy}: & n=8 & 1014 & 684 & 810 & 990 & 840 & 978 & 1002 & 1110 & & \end{array}$$

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

$$\begin{aligned} \text{Wilcoxon} \quad W &= R_2 = 101 \\ \text{Mann-Whitney} \quad U &= R_2 - \frac{n_2(n_2+1)}{2} = 101 - 36 = 65 \end{aligned}$$

- 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

Correction

$$\begin{array}{rlll} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a & : \eta_1 > \eta_2 & \text{as} & W = R_2 > T_L \\ \textbf{Reject } H_0 \text{ against } H_a & : \eta_1 < \eta_2 & \text{as} & W = R_2 > T_U \\ \textbf{Reject } H_0 \text{ against } H_a & : \eta_1 \neq \eta_2 & \text{as} & W = R_2 > T_U \end{array}$$

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

Correction

$$\begin{array}{rlll} \textbf{Do not} \text{ reject } H_0 \text{ against } H_a & : \eta_1 > \eta_2 & \text{as} & Z > C_R = -1.645 \\ \textbf{Reject } H_0 \text{ against } H_a & : \eta_1 < \eta_2 & \text{as} & Z > C_R = 1.645 \\ \textbf{Reject } H_0 \text{ against } H_a & : \eta_1 \neq \eta_2 & \text{as} & Z > C_{R_2} = 1.960 \end{array}$$

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