2. One-sample Location Problem

Location problem

- A class of distributions, denoted by $\{f_a(x), a \in \mathbb{R}\}$, is said to be a *location family* if each $f_a(x)$ in the class is a probability function (pf) of the form

$$f_a(x) = f(x-a) \text{ for some common pf } f(x)$$

- The real value a is called the *location parameter* of the family.
- Equivalently, each member of a location family has a cdf of the form

$$F_a(x) = F(x-a) \text{ for some common cdf } F(x)$$

- For example, the normal distributions $N(\mu, \sigma^2)$ with fixed $\sigma > 0$ form a location family with location parameter $\mu \in \mathbb{R}$ and $f(x)$ being the density of the $N(0, \sigma^2)$ distribution.

- An increase of the location parameter a shifts the distribution of $X \sim f(x-a)$ to the right without changing its shape, which increases the value of X in a probabilistic or stochastic sense.

- More specifically, let $X \sim f(x-a)$, $Y \sim f(x-b)$. If $a < b$, then

$$x-a > x-b \;\Rightarrow\; \Pr(X \le x) = F(x-a) \ge F(x-b) = \Pr(Y \le x) \text{ for all } x \in \mathbb{R}$$

  and $\Pr(X \le x) > \Pr(Y \le x)$ for some $x \in \mathbb{R}$.

  This can be interpreted as "X is more likely to take small values than Y".

- In the above sense and according to the definition of *stochastic order*, we say that "X is less than Y in stochastic order", written $X <_{\mathrm{st}} Y$.

- Thus the location parameter represents the magnitude of a random variable in a stochastic sense. The mean and median of a random variable have been commonly taken as the location parameters in practice.

- Statistical inference on one location parameter from one sample is referred to as the *one-sample location problem*. It will be tackled by hypothesis tests and estimation of the median in this course.

- We start with the *sign test* of the median.

2.1 Sign test

Let $X_1,\ldots,X_n$ represent a sample of random variables of size n.

Assumption 2.1

(i) $X_1,\ldots,X_n$ are mutually independent;

(ii) $X_1,\ldots,X_n$ are continuous with a common median θ (not necessarily have the same distribution).

Null hypothesis: $H_0:\theta=0$.

Test statistic: $B=\sum_{i=1}^{n} I_{\{X_i>0\}}$ = the number of positive X_i's.

Under $H_0:\theta=0$, $\Pr(X_i>0)=\Pr(X_i>\theta)=0.5 \Rightarrow B \sim Bin(n,0.5)$.

From the binomial distribution, the mean and variance of B under H_0 are given respectively by

$$\mathrm{E}_0[B]=0.5n \quad \text{and} \quad \mathrm{Var}_0(B)=0.5(1-0.5)n=0.25n$$

Rejection rule: Reject H_0 at (achievable) level α of significance if

- $B \geq b_\alpha$ against $H_1: \theta > 0$ (one-sided upper-tail test);
- $B \leq n - b_\alpha$ against $H_1: \theta < 0$ (one-sided lower-tail test);
- either $B \geq b_{\alpha/2}$ or $B \leq n - b_{\alpha/2}$ against $H_1: \theta \neq 0$ (two-sided test),

where $b_\alpha \in \{0,1,\ldots,n\}$ satisfies $\Pr(B \geq b_\alpha) = \alpha$ for $B \sim Bin(n, 0.5)$.

Approximate rejection rule: Reject H_0 at the α level of significance if

- $B^* \geq z_\alpha$ against $H_1: \theta > 0$;
- $B^* \leq -z_\alpha$ against $H_1: \theta < 0$;
- $|B^*| \geq z_{\alpha/2}$ against $H_1: \theta \neq 0$,

where $\Pr(Z \geq z_\alpha) = \alpha$ for $Z \sim N(0,1)$ and

$$B^* = \frac{B - \mathrm{E}_0[B]}{\sqrt{\mathrm{Var}_0(B)}} = \frac{B - 0.5n}{0.5\sqrt{n}}$$

Remark 2.1

- When $X_1,\ldots,X_n$ are continuous, $\Pr(X_i=0)=0$. In practice, however, it is possible to observe zeros from X_i's (due to rounding or some other reasons). In that case, a sensible option is to discard those zero values and replace the sample size n by the number of nonzero observations.

- If we wish to test $H_0:\theta=\theta_0$ for a specified value $\theta_0\neq 0$, we can replace each X_i with $X_i-\theta_0$, $i=1,\ldots,n$. Then the test of $H_0:\theta=\theta_0$ based on $X_1,\ldots,X_n$ is equivalent to the test of $H_0:\theta=0$ based on $X_1-\theta_0,\ldots,X_n-\theta_0$.

- The sign test has the advantage of only requiring independence and a common median, hence is applicable to test the median in most situations.

- On the other hand, the sign test only utilizes the signs of the data, but not their values, hence it is considered as less efficient for underuse of information from the data. The *Wilcoxon signed rank test* introduced in Subsection 2.2 next is more efficient when it also uses the magnitude of the data.

Paired data

- In practice, we are often interested in comparing two measures, say X and Y, from the same subject. A typical example is to compare certain medical measurements of the same patient before and after a treatment.

- Let $Z = Y - X$ be a continuous random variable with median θ, and its pdf $f(z) > 0$ in a neighbourhood of θ. Then
  - $\theta = 0 \;\Rightarrow\; \Pr(Z > 0) = \Pr(Z < 0) = 0.5 \;\Rightarrow\; \Pr(Y > X) = \Pr(Y < X)$; and
  - $\theta > 0 \;\Rightarrow\; \Pr(Z > 0) > \Pr(Z > \theta) = 0.5 = \Pr(Z < \theta) > \Pr(Z < 0)$

    $\Rightarrow\; \Pr(Y > X) > \Pr(Y < X)$ (Y is more likely greater than X).

- If X and Y are the measurements from a patient before and after a treatment, and a larger value is better (such as physical strength), then $\theta > 0$ indicates a positive effect of the treatment.

- In such a case, an acceptance of $H_0 : \theta = 0$ by a hypothesis test means that the treatment is ineffective, and the rejection of H_0 in favor of $H_1 : \theta > 0$ can be interpreted as the evidence for the positive effect of the treatment.

- Similarly, if a smaller measurement indicates an improvement, then a test of H_0 against $H_1: \theta < 0$ helps to determine the effectiveness of the treatment.
- The two-sided test of H_0 against $H_1: \theta \neq 0$ is used to detect any effect of the treatment – either positive or negative.
- The parameter θ is often referred to as the *treatment effect*.
- Let X_i and Y_i be the measurements on subject *i* and $Z_i = Y_i - X_i$, $i = 1, \ldots, n$. Then a sign test based on observed data from $Z_1, \ldots, Z_n$ can be applied to test the hypotheses on the treatment effect θ.
- $(X_1, Y_1), \ldots, (X_n, Y_n)$ are referred to as *paired data*, or *matched pairs*.
- Statistical inference on paired data is considered as a one-sample problem, since it is drawn on the differences $Z_1, \ldots, Z_n$ between the pairs, although two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are involved.
- Independence is assumed between the differences $Z_1, \ldots, Z_n$, but not needed within each pair (X_i, Y_i).

Estimation of θ associated with the sign test

- Let $X_{(1)} \le \cdots \le X_{(n)}$ be the order statistics of $X_1,\ldots,X_n$. A nonparametric estimator for the median (treatment effect) θ is given by

$$\tilde{\theta} = \text{median}\{X_i, 1 \le i \le n\} = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd;} \\ \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2} & \text{if } n \text{ is even.} \end{cases}$$

- Let $B_\theta =$ number of $\{i : X_i \ge \theta\}$. Then $\Pr(X_i \ge \theta) = 0.5 \;\Rightarrow\; B_\theta \sim Bin(n, 0.5)$.
- Since $X_{(k)} < \theta \;\Leftrightarrow\; \#\{i : X_i < \theta\} \ge k \;\Leftrightarrow\; \#\{i : X_i \ge \theta\} \le n-k \;\Leftrightarrow\; B_\theta \le n-k$, $\Pr(X_{(k)} < \theta) = \Pr(B_\theta \le n-k) = \Pr(B_\theta \ge k)$. Consequently,

$$\Pr(X_{(b_{\alpha/2})} \le \theta) = \Pr(X_{(b_{\alpha/2})} < \theta) = \Pr(B_\theta \ge b_{\alpha/2}) = \alpha/2 \qquad (2.1)$$

- Let $C_\alpha = n+1-b_{\alpha/2}$. Then by the symmetry of B_θ,

$$\Pr(X_{(C_\alpha)} < \theta) = \Pr(B_\theta \ge n+1-b_{\alpha/2}) = \Pr(B_\theta \le b_{\alpha/2}-1) = \Pr(B_\theta < b_{\alpha/2})$$

- It follows that

$$\Pr\left(X_{(C_\alpha)} \geq \theta\right) = 1 - \Pr\left(B_\theta < b_{\alpha/2}\right) = \Pr\left(B_\theta \geq b_{\alpha/2}\right) = \alpha/2 \tag{2.2}$$

- Let $\theta_L = X_{(C_\alpha)}$ and $\theta_U = X_{(n+1-C_\alpha)} = X_{(b_{\alpha/2})}$. Then (2.1) and (2.2) imply

$$\Pr(\theta_L < \theta < \theta_U) = 1 - \Pr\left(X_{(C_\alpha)} \geq \theta\right) - \Pr\left(X_{(b_{\alpha/2})} \leq \theta\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha$$

- Thus a $100(1-\alpha)\%$ confidence interval for θ is given by

$$(\theta_L, \theta_U) = \left(X_{(C_\alpha)}, X_{(n+1-C_\alpha)}\right) = \left(X_{(n+1-b_{\alpha/2})}, X_{(b_{\alpha/2})}\right) \tag{2.3}$$

  The $1-\alpha$ in (2.3) should be achievable and not below the desired level.

- For large n, C_α can be approximated by

$$C_\alpha \approx \mathrm{E}_0[B] - z_{\alpha/2}\sqrt{\mathrm{Var}_0(B)} = 0.5n - z_{\alpha/2} 0.5\sqrt{n} \tag{2.4}$$

- We may round the value on the right side of (2.4) to the nearest integer, or more conservatively, take its integer part.

Example 2.1 In Example 3.5 from page 65 of the textbook, paired data (X_i, Y_i) are provided on average *beak-clapping counts* (per minute) in 25 white leghorn chick embryos, where X_i and Y_i represent the counts during the dark period and illumination period for embryo i, $i = 1, \ldots, 25$.

The question of interest is whether there is strong evidence that light stimulus increases the responsivity of embryos (measured by beak-clapping counts).

The data are presented in Table 3.5 on page 66, which also includes $Z_i = Y_i - X_i$ and $\psi_i = I_{\{Z_i > 0\}}$, $i = 1, \ldots, 25$. The values of $Z_1, \ldots, Z_{25}$ are copied below:

i	Z_i	i	Z_i	i	Z_i	i	Z_i	i	Z_i
1	-0.8	6	54.0	11	-8.5	16	20.6	21	4.7
2	7.5	7	48.3	12	7.1	17	25.0	22	24.7
3	46.9	8	3.9	13	40.7	18	24.7	23	52.8
4	17.6	9	16.7	14	23.8	19	-1.8	24	8.5
5	-4.6	10	19.7	15	14.8	20	21.9	25	1.9

$B = 21$ positive Z_i values are observed.

Since $Z > 0$ points to higher responsivity of embryos under light stimulus, a sign test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ based on $Z_1, \ldots, Z_n$ can be applied to determine whether light stimulus increases responsivity. For $B \sim Bin(25, 0.5)$,

$$\Pr(B \geq 18) = 0.022 \quad \text{and} \quad \Pr(B \geq 17) = 0.054 \tag{2.5}$$

Thus, to achieve the 5% level, take $\alpha = 0.022$ and $b_\alpha = b_{0.022} = 18$. Then the sign test rejects $H_0 : \theta = 0$ in favor of $H_1 : \theta > 0$ if $B \geq b_{0.022} = 18$.

With $B = 21 > 18$, there is sufficient evidence at the 5% level of significance to reject $H_0 : \theta = 0$ and conclude $\theta > 0$ (light stimulus increases responsivity).

The *p*-value (interpreted as the *achieved level*) is $\Pr(B \geq 21) = 0.00046 < 0.001$, which shows overwhelming evidence for $\theta > 0$.

The approximate rule to reject H_0 at the 5% level is $B^* \geq z_{0.05} = 1.645$. Hence

$$B^* = \frac{B - 0.5n}{0.5\sqrt{n}} = \frac{21 - 12.5}{2.5} = 3.40 > 1.645 \;\Rightarrow\; \text{Reject } H_0 \;\Rightarrow\; \theta > 0$$

The approximate *p*-value is $\Pr(B^* \geq 3.40) \approx 0.00034$.

To estimate the treatment effect θ, order $Z_1, \ldots, Z_{25}$ as $Z_{(1)} < \cdots < Z_{(25)}$:

i	$Z_{(i)}$	i	$Z_{(i)}$	i	$Z_{(i)}$	i	$Z_{(i)}$	i	$Z_{(i)}$
1	-8.5	6	3.9	11	14.8	16	21.9	21	40.7
2	-4.6	7	4.7	12	16.7	17	23.8	22	46.9
3	-1.8	8	7.1	13	17.6	18	24.7	23	48.3
4	-0.8	9	7.5	14	19.7	19	24.7	24	52.8
5	1.9	10	8.5	15	20.6	20	25.0	25	54.0

Then the treatment effect θ is estimated by $\tilde{\theta} = Z_{((25+1)/2)} = Z_{(13)} = 17.6$.

Take $\alpha/2 = 0.022 \Rightarrow \alpha = 0.044$. Then by (2.3) and (2.5), a $100(1-\alpha)\% = 95.6\%$ confidence interval of θ is given by

$$(\theta_L, \theta_U) = \left(Z_{(25+1-b_{0.022})}, Z_{(b_{0.022})}\right) = \left(Z_{(8)}, Z_{(18)}\right) = (7.1,\ 24.7)$$

By (2.4), $C_{0.05} \approx 0.5 \times 25 - 1.96 \times 0.5 \times 5 = 7.6$. Round it to 8 to get $\left(Z_{(8)}, Z_{(18)}\right)$, or take integer part 7 to get a more conservative interval $\left(Z_{(7)}, Z_{(19)}\right) = (4.7,\ 24.7)$.

More details of this example can be found in Examples 3.5 – 3.7 of the textbook.

2.2 Wilcoxon signed rank test

Assumption 2.2

(i) The sample random variables $X_1,\ldots,X_n$ are independent;

(ii) The probability distributions of $X_1,\ldots,X_n$ are continuous and symmetric about a common median θ (not necessarily identical).

Null hypothesis: $H_0:\theta=0$, where θ is the median of $X_1,\ldots,X_n$.

Rank: Assume no ties among $|X_1|,\ldots,|X_n|$. Let $|X|_{(1)}<\cdots<|X|_{(n)}$ be ordered values of $|X_1|,\ldots,|X_n|$. Define the *rank* R_i of X_i by $R_i=k$ if $|X_i|=|X|_{(k)}$. That is, the X_i with the k^{th} smallest absolute value has rank $R_i=k$.

Example 2.2 $(X_1,\ldots,X_6)=(-3,4,1,7,-9,5)\ \Rightarrow\ \left(|X_1|,\ldots,|X_6|\right)=(3,4,1,7,9,5)\ \Rightarrow$

$\left(|X|_{(1)},\ldots,|X|_{(6)}\right)=(1,3,4,5,7,9)=\left(|X_3|,|X_1|,|X_2|,|X_6|,|X_4|,|X_5|\right)\ \Rightarrow$

$\left(R_1,R_2,R_3,R_4,R_5,R_6\right)=(2,3,1,5,6,4)$

Test statistic: There are several equivalent forms of the Wilcoxon signed rank test statistic. We will consider the following form:

$$T^+ = \sum_{i=1}^{n} R_i \psi_i, \quad \text{where } \psi_i = I_{\{X_i > 0\}},\ i = 1, \dots, n. \tag{2.6}$$

For $(X_1, \dots, X_6) = (-3, 4, 1, 7, -9, 5)$ in Example 2.2, $\left(R_1, \dots, R_6\right) = (2, 3, 1, 5, 6, 4)$ and $(\psi_1, \dots, \psi_6) = (0, 1, 1, 1, 0, 1)$. Hence

$$T^+ = R_1\psi_1 + \cdots + R_6\psi_6 = R_2 + R_3 + R_4 + R_6 = 3 + 1 + 5 + 4 = 13$$

It would be more convenient to calculate T^+ by rearranging $(X_1, \dots, X_n)$ such that $|X_1| < \cdots < |X_n|$. Then $\left(R_1, \dots, R_n\right) = (1, \dots, n)$ and

$$T^+ = \sum_{i=1}^{n} R_i \psi_i = \sum_{i=1}^{n} i\psi_i = \psi_1 + 2\psi_2 + \cdots + n\psi_n$$

In Example 2.2, rearrange $(X_1, \dots, X_6) = (-3, 4, 1, 7, -9, 5)$ to $(1, -3, 4, 5, 7, -9)$. Then $(\psi_1, \dots, \psi_6) = (1, 0, 1, 1, 1, 0)$ and $T^+ = \psi_1 + 3\psi_3 + 4\psi_4 + 5\psi_5 = 1 + 3 + 4 + 5 = 13$.

Exact distribution of T^+

Since θ is the median of X_i, $\Pr(\psi_i = 1) = \Pr(X_i > 0) = \Pr(X_i > \theta) = 0.5$ under H_0. Consider a statistical experiment as follows:

- Draw an ordered $(\psi_1, \ldots, \psi_n)$ randomly from $\{0,1\}$ with replacement, with $\Pr(\psi_i = 1) = \Pr(\psi_i = 0) = 0.5$, $i = 1, \ldots, n$.
- Define the outcome as $\omega = (\psi_1, 2\psi_2, \ldots, n\psi_n)$. The sample space Ω consists of 2^n equally likely outcomes determined by n-tuples $(\psi_1, \ldots, \psi_n)$.
- Each outcome ω can be equivalently expressed as $\omega = (r_1, \ldots, r_B)$, where B is the number of 1's in $(\psi_1, \ldots, \psi_n)$ and $r_1 < \cdots < r_B \in \{1, 2, \ldots, n\}$ if $B > 0$ or $\omega = \phi$ (empty set) if $B = 0$ with $(\psi_1, \ldots, \psi_n) = (0, \ldots, 0)$.
- For example, if $n = 8$ and $(\psi_1, \ldots, \psi_n) = (\psi_1, \ldots, \psi_8) = (0,1,1,0,1,0,1,1)$, then $\omega = (\psi_1, 2\psi_2, \ldots, 8\psi_8) = (0,2,3,0,5,0,7,8)$ can be equivalently expressed by $\omega = (r_1, r_2, r_3, r_4, r_5) = (2,3,5,7,8)$ with $B = 5$.

Let "~" denote "have the same distribution as". Then under Assumption 2.2, the distribution of T^+ in (2.6) can be determined by

$$T^+ = \sum_{i=1}^{n} R_i \psi_i \sim \sum_{i=1}^{n} i\psi_i = \sum_{i=1}^{B} r_i \tag{2.7}$$

See Appendix for the proof of this equivalence in distribution.

The range of T^+ is $\{0,1,\ldots,M\}$ with $M = 1+2+\cdots+n = n(n+1)/2$. By (2.7) and equally likely $\omega \in \Omega$, the distribution of T^+ under H_0 is given by

$$\Pr(T^+ = t) = \frac{\text{Number of } \omega = (r_1,\ldots,r_B) : r_1+\cdots+r_B = t}{2^n} \tag{2.8}$$

for $t \in \{0,1,\ldots,M\}$, with $T^+ = 0$ if and only if $B = 0$ $(\omega = \phi)$.

Note: Given $X_1,\ldots,X_n$, we can calculate $T^+ = r_1+\cdots+r_B$, such as $B = 4$ and $T^+ = r_1+r_2+r_3+r_4 = 1+3+4+5 = 13$ in Example 2.2. However, (2.7) – (2.8) are valid only for symmetric $X_1,\ldots,X_n$. Without this assumption, $T^+ \sim r_1+\cdots+r_B$ does not necessarily hold, which would invalidate (2.7) – (2.8).

Example 2.3 Let $n = 3$. Then

$$M = 1+2+3 = \frac{3(3+1)}{2} = 6 \quad \text{and} \quad 2^n = 2^3 = 8$$

The distribution of T^+ is calculated by (2.8) in the following table.

t	(ψ_1, ψ_2, ψ_3)	$(\psi_1, 2\psi_2, 3\psi_3)$	B	$(r_1, \ldots, r_B)$	$\Pr(T^+ = t)$
0	$(0,0,0)$	$(0,0,0)$	0	ϕ	$1/8 = 0.125$
1	$(1,0,0)$	$(1,0,0)$	1	$(r_1) = (1)$	$1/8 = 0.125$
2	$(0,1,0)$	$(0,2,0)$	1	$(r_1) = (2)$	$1/8 = 0.125$
3	$(0,0,1)$ $(1,1,0)$	$(0,0,3)$ $(1,2,0)$	1 2	$(r_1) = (3)$ $(r_1, r_2) = (1,2)$	$2/8 = 0.25$
4	$(1,0,1)$	$(1,0,3)$	2	$(r_1, r_2) = (1,3)$	$1/8 = 0.125$
5	$(0,1,1)$	$(0,2,3)$	2	$(r_1, r_2) = (2,3)$	$1/8 = 0.125$
6	$(1,1,1)$	$(1,2,3)$	3	$(r_1, r_2, r_3) = (1,2,3)$	$1/8 = 0.125$

Approximate distribution of T^+

From (2.7), it is easy to calculate the mean and variance of T^+ under $H_0 : \theta = 0$:

$$E_0[\psi_i] = \Pr_0(\psi_i = 1) = 0.5 \;\Rightarrow$$

$$E_0[T^+] = E_0\left[\sum_{i=1}^{n} i\psi_i\right] = \sum_{i=1}^{n} iE_0[\psi_i] = \frac{1}{2}\sum_{i=1}^{n} i = \frac{1}{2}\cdot\frac{n(n+1)}{2} = \frac{n(n+1)}{4} \tag{2.9}$$

$$\mathrm{Var}_0(\psi_i) = 0.5(1-0.5) = 0.25 \;\Rightarrow$$

$$\mathrm{Var}_0(T^+) = \mathrm{Var}_0\left(\sum_{i=1}^{n} i\psi_i\right) = \sum_{i=1}^{n} i^2\mathrm{Var}_0(\psi_i) = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{24} \tag{2.10}$$

Then by the central limit theorem,

$$T^* = \frac{T^+ - E_0[T^+]}{\sqrt{\mathrm{Var}_0(T^+)}} = \frac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1) \tag{2.11}$$

approximately for large n under H_0.

Rejection rule: Let $M = n(n+1)/2$ and $\Pr(T^+ \ge t_\alpha) = \alpha$ with $t_\alpha \in \{0,1,\ldots,M\}$ under H_0. Then the Wilcoxon signed rank test rejects H_0 at the α level if

- $T^+ \ge t_\alpha$ against $H_1 : \theta > 0$;
- $T^+ \le M - t_\alpha$ against $H_1 : \theta < 0$;
- either $T^+ \ge t_{\alpha/2}$ or $T^+ \le M - t_{\alpha/2}$ against $H_1 : \theta \ne 0$.

The level α is achievable if $\Pr(T^+ \ge t) = \alpha$ for some $t \in \{0,1,\ldots,M\}$.

The p-value based on the observed value t^+ of T^+ is $\Pr(T^+ \ge t^+)$, $\Pr(T^+ \le t^+)$ and $2\Pr(T^+ \ge t^+)$ against $H_1 : \theta > 0$, $\theta < 0$ and $\theta \ne 0$, respectively.

In Example 2.3, the achievable level α includes

$$\Pr(T^+ \ge 6) = 0.125,\ \ \Pr(T^+ \ge 5) = 0.25,\ \ \Pr(T^+ \ge 4) = 0.375, \ldots$$

with $t_{0.125} = 6$, $t_{0.25} = 5, \ldots$. If $T^+ = 6$ is observed, then the p-value for $\theta > 0$ is $\Pr(T^+ \ge 6) = 0.125$. In this example, H_0 is never rejected if $\alpha < 0.125$.

Approximate rejection rule:

Let T^* be defined in (2.11). Since $T^* \sim N(0,1)$ approximately, the approximate rejection rules for $H_0: \theta = 0$ are given as follows:

Reject H_0 at the α level if

- $T^* \geq z_\alpha$ against $H_1: \theta > 0$, or equivalently,

$$T^+ \geq \mathrm{E}_0[T^+] + z_\alpha \sqrt{\mathrm{Var}_0(T^+)}$$

  with $\mathrm{E}_0[T^+]$ and $\mathrm{Var}_0(T^+)$ given by (2.9) and (2.10), respectively;
- $T^* \leq -z_\alpha$ against $H_1: \theta < 0$, or $T^+ \leq \mathrm{E}_0[T^+] - z_\alpha \sqrt{\mathrm{Var}_0(T^+)}$;
- $|T^*| \geq z_{\alpha/2}$ against $H_1: \theta \neq 0$.

The approximate rule is often used in practice, since the exact distribution of T^+ is time-consuming to calculate for large or even moderate n, while the normal approximation is good with reasonable n due to the symmetry of T^+. But we can also use R to calculate the exact distribution of T^+.

Example 2.4 Consider the case of $n=6$ and $H_1:\theta>0$ at the desired level 5%. Then $2^n=2^6=64$ and $M=6(6+1)/2=42=21$. Hence $\Pr(T^+\geq 19)=\Pr(T^+\leq 2)$. Since there are 3 outcomes such that $T^+\leq 2$: $\omega=\phi$ for $T^+=0$, $\omega=(1)$ for $T^+=1$ and $\omega=(2)$ for $T^+=2$, we obtain

$$\Pr(T^+\geq 19)=\Pr(T^+\leq 2)=3/2^n=3/64=0.047<0.05$$

Similarly, there are two outcomes for $T^+=3$: $\omega=(3)$ and $\omega=(1,2)$, hence

$$\Pr(T^+\geq 18)=\Pr(T^+\leq 3)=\Pr(T^+\leq 2)+\Pr(T^+=3)=(3+2)/64=0.078>0.05$$

Thus the exact rule to reject H_0 is $T^+\geq t_{0.047}=19$. On the other hand,

$$\mathrm{E}_0[T^+]+z_{0.05}\sqrt{\mathrm{Var}_0(T^+)}=\frac{6(6+1)}{4}+1.645\sqrt{\frac{6(6+1)(12+1)}{24}}=18.35$$

This shows that the approximate rejection rule $T^+\geq 18.35$ for $H_1:\theta>0$ at the 5% level is the same as the exact rule $T^+\geq 19$ at the 4.7% level.

Ties: Under the assumptions of independent and continuous $X_1, \ldots, X_n$, there is zero probability of any ties. In practice, however, ties may occur due to rounding or other reasons. In that case, we assign the average rank to tied values.

For example, if $|X_1| < |X_2| = |X_3| < |X_4|$, then $R_1 = 1$, $R_2 = R_3 = 2.5$ and $R_4 = 4$; if $|X_1| = |X_2| = |X_3| < |X_4|$, then $R_1 = R_2 = R_3 = 2$ and $R_4 = 4$.

In such a case, the distribution of T^+ conditional on ties differs from that with no ties. This does not affect $\mathrm{E}_0[T^+]$ in (2.9), but the variance in (2.10) reduces to

$$\mathrm{Var}_0(T^+) = \frac{n(n+1)(2n+1)}{24} - \frac{1}{48}\sum_{j=1}^{g} t_j(t_j - 1)(t_j + 1), \qquad (2.12)$$

where g is the number of groups with tied ranks, and t_j is the number of tied ranks in group j, $j = 1, \ldots, g$. It does not matter whether we count an untied rank as a group with $t_j = 1$, or ignore it, since $t_j(t_j - 1)(t_j + 1) = 0$ when $t_j = 1$.

In particular, if $t_j = 1$ for $j = 1, \ldots, k$ (no ties), then (2.12) reduces to (2.10).

To justify (2.12), suppose that t points in the sample are tied in absolute value: $|X|_{(k)} = |X|_{(k+1)} = \cdots = |X|_{(k+t-1)}$. Then each of them is assigned the average rank

$$\frac{1}{t}\sum_{j=0}^{t-1}(k+j) = \frac{1}{t}\sum_{j=0}^{t-1}k + \frac{1}{t}\sum_{j=0}^{t-1}j = \frac{1}{t}\left[kt + \frac{t(t-1)}{2}\right] = k + \frac{t-1}{2}$$

Thus the sum of squares of their ranks is

$$t\left(k+\frac{t-1}{2}\right)^2 = t\left[k^2 + (t-1)k + \left(\frac{t-1}{2}\right)^2\right] = tk^2 + t(t-1)k + \frac{t(t-1)^2}{4}$$

If there were no ties, then these t values would occupy ranks $k, k+1, \ldots, k+t-1$, with the sum of squares:

$$\sum_{j=0}^{t-1}(k+j)^2 = \sum_{j=0}^{t-1}\left(k^2 + 2kj + j^2\right) = k^2 t + 2k\sum_{j=1}^{t-1} j + \sum_{j=1}^{t-1} j^2$$

$$= tk^2 + kt(t-1) + \frac{t(t-1)(2t-1)}{6}$$

The difference between these two sums of squares ("no ties" minus "ties") is

$$\sum_{j=0}^{t-1}(k+j)^2 - t\left(k+\frac{t-1}{2}\right)^2 = \frac{t(t-1)(2t-1)}{6} - \frac{t(t-1)^2}{4}$$

$$= \frac{t(t-1)\left[2(2t-1)-3(t-1)\right]}{12}$$

$$= \frac{t(t-1)\left[4t-2-3t+3\right]}{12} = \frac{t(t-1)(t+1)}{12}$$

Compare with (2.10), we can see that the variance with no ties minus the variance with one group of t tied values is

$$\frac{1}{4}\left[\sum_{j=0}^{t-1}(k+j)^2 - t\left(k+\frac{t-1}{2}\right)^2\right] = \frac{t(t-1)(t+1)}{48}$$

This explains (2.12).

Example 2.5 Consider two cases with $n = 4$:

(i) $0 < |X_1| < |X_2| < |X_3| < |X_4|$ (no ties)

(ii) $0 < |X_1| < |X_2| < |X_3| = |X_4|$ (tied $R_3 = R_4 = 3.5$)

The distributions of T^+ under H_0 for the two cases are shown below:

(i) $(R_1, R_2, R_3, R_4) = (1,2,3,4)$			(ii) $(R_1, R_2, R_3, R_4) = (1,2,3.5,3.5)$		
t	$(r_1,\ldots,r_B)$	$\Pr(T^+ = t)$	t	$(r_1,\ldots,r_B)$	$\Pr(T^+ = t)$
0	ϕ	1/16	0	ϕ	1/16
1	(1)	1/16	1	(1)	1/16
2	(2)	1/16	2	(2)	1/16
3	(3), (1,2)	2/16	3	(1,2)	1/16
4	(4), (1,3)	2/16	3.5	(3.5), (3.5)	2/16
5	(1,4), (2,3)	2/16	4.5	(1,3.5), (1,3.5)	2/16
$\Pr(T^+ = t) = \Pr(T^+ = 10 - t)$ for $t = 6,7,8,9,10.$			$\Pr(T^+ = t) = \Pr(T^+ = 10 - t)$ for $t = 5.5, 6.5, 7, 8, 9, 10.$		

In Case (i), the mean and variance of T^+ under H_0 are

$$\mathrm{E}_0[T^+] = \frac{4(4+1)}{4} = 5 \quad \text{and} \quad \mathrm{Var}_0(T^+) = \frac{4(4+1)(2\times 4+1)}{24} = \frac{15}{2} = 7.5$$

In Case (ii),

$$\mathrm{E}_0[T^+] = \frac{1+2+3+7+8+9+10+2(3.5+4.5+5.5+6.5)}{16} = \frac{40+2(20)}{16} = 5$$

(same as with no ties), and conditional on $R_3 = R_4 = 3.5$,

$$\mathrm{Var}_0(T^+) = \frac{1+2^2+3^2+7^2+8^2+9^2+10^2+2(3.5^2+4.5^2+5.5^2+6.5^2)}{16} - 5^2$$

$$= 32.375 - 25 = 7.375 \quad (< 7.5 \text{ with no ties})$$

If we use (2.12) with $g = 1$ and $t_1 = 2$, then

$$\mathrm{Var}_0(T^+) = 7.5 - \frac{2(2-1)(2+1)}{48} = 7.5 - 0.125 = 7.375$$

This matches the result from the direct calculation using the distribution of T^+.

Symmetry of T^+: From Examples 2.3 and 2.5, we can see that the distribution of T^+ under H_0 is symmetric about its mean, even if conditional on ties. In fact, this is true for any n.

To see why, note that $\psi_i \sim 1-\psi_i \sim Bin(1,0.5)$, hence

$$T^+ = \sum_{i=1}^{n}\psi_i R_i \sim \sum_{i=1}^{n}(1-\psi_i)R_i = \sum_{i=1}^{n}R_i - \sum_{i=1}^{n}\psi_i R_i = \frac{n(n+1)}{2} - T^+ = M - T^+ \quad (2.13)$$

As a result, under H_0,

$$\Pr(T^+ \le M-t) = \Pr(M-T^+ \ge t) = \Pr(T^+ \ge t) \quad \text{for all } t \in \{0,1,2,\ldots,M\}$$

Thus T^+ is symmetric about its mean $\mathrm{E}_0[T^+] = M/2 = n(n+1)/4$ (recall equation (1.4) for symmetric distributions in Section 1).

It also justifies the exact rejection rule $T^+ \le M - t_\alpha$ against $H_1: \theta < 0$ since

$$\Pr(T^+ \le M - t_\alpha) = \Pr(M - T^+ \ge t_\alpha) = \Pr(T^+ \ge t_\alpha) = \alpha$$

Equivalent versions

Let $T^- = \sum_i (1-\psi_i)R_i$ be the sum of the ranks of negative X_i's. Then

$$T^+ + T^- = \sum_{i=1}^n \psi_i R_i + \sum_{i=1}^n (1-\psi_i)R_i = \sum_{i=1}^n R_i = 1+2+\cdots+n = M$$

This shows that $T^- = M - T^+$ and hence T^- is an equivalent test statistic to T^+. The rejection rule $T^+ \geq t_\alpha$ for $H_1 : \theta > 0$ is equivalent to $T^- \leq M - t_\alpha$.

Another equivalent version of the signed rank test statistic is

$$W = \sum_{i=1}^n \mathrm{sgn}(X_i)R_i = \sum_{i=1}^n \left(I_{\{X_i>0\}} - I_{\{X_i<0\}}\right)R_i = T^+ - T^- = 2T^+ - M,$$

where

$$\mathrm{sgn}(x) = I_{\{x>0\}} - I_{\{x<0\}} = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Thus T^+ does not lose information for relying on the ranks of positive X_i's.

Estimation of the median

- Let $W_{(1)} \le W_{(2)} \le \cdots \le W_{(M)}$ be ordered values of the $M = n(n+1)/2$ averages of (X_i, X_j) (known as the *Walsh averages*):

$$W_{ij} = \frac{X_i + X_j}{2}, \quad i \le j = 1, \ldots, n. \tag{2.14}$$

- The median (treatment effect) θ can be estimated by

$$\hat{\theta} = \text{median}\left\{\frac{X_i + X_j}{2}, i \le j = 1, \ldots, n\right\}$$

$$= \begin{cases} W_{((M+1)/2)} & \text{if } M \text{ is odd;} \\ \dfrac{W_{(M/2)} + W_{(M/2+1)}}{2} & \text{if } M \text{ is even.} \end{cases} \tag{2.15}$$

- A confidence interval for the median θ can be obtained based on the ordered Walsh averages $W_{(1)} \le W_{(2)} \le \cdots \le W_{(M)}$ as follows.

It is proved in Appendix that $W_{(k)} < 0 < W_{(k+1)} \Leftrightarrow T^+ = M - k$. Let $\Pr_0$ denote the probability with $\theta = 0$. Then $\Pr(W_{(k)} - \theta \le x) = \Pr_0(W_{(k)} \le x)$ for all x since the median of $X_i - \theta$ is 0. Define $W_{(0)} = -\infty$ and $W_{(M+1)} = \infty$. Then

$$\Pr\left(\theta < W_{(k)}\right) = \sum_{l=0}^{k-1} \Pr\left(W_{(l)} < \theta < W_{(l+1)}\right) = \sum_{l=0}^{k-1} \Pr_0\left(W_{(l)} < 0 < W_{(l+1)}\right)$$

$$= \sum_{l=0}^{k-1} \Pr_0\left(T^+ = M - l\right) = \Pr_0\left(T^+ \ge M - k + 1\right) \qquad (2.16)$$

Define $C_\alpha = M - t_{\alpha/2} + 1$. It follows from (2.16) that

$$\Pr\left(\theta < W_{(C_\alpha)}\right) = \Pr_0\left(T^+ \ge M - C_\alpha + 1\right) = \Pr_0\left(T^+ \ge t_{\alpha/2}\right) = \frac{\alpha}{2}$$

and

$$\Pr\left(\theta > W_{(t_{\alpha/2})}\right) = 1 - \Pr\left(\theta < W_{(t_{\alpha/2})}\right) = 1 - \Pr_0\left(T^+ \ge M - t_{\alpha/2} + 1\right)$$

$$= \Pr_0\left(T^+ < M - t_{\alpha/2} + 1\right) = \Pr_0\left(T^+ \le M - t_{\alpha/2}\right) = \frac{\alpha}{2}$$

Consequently,

$$\Pr\left(W_{(C_\alpha)} < \theta < W_{(t_{\alpha/2})}\right) = 1 - \Pr\left(\theta < W_{(C_\alpha)}\right) - \Pr\left(\theta > W_{(t_{\alpha/2})}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

Thus a $100(1-\alpha)\%$ confidence interval for θ is given by

$$(\theta_L, \theta_U) = \left(W_{(C_\alpha)}, W_{(M+1-C_\alpha)}\right) = \left(W_{(M+1-t_{\alpha/2})}, W_{(t_{\alpha/2})}\right) \tag{2.17}$$

For large n, because $C_\alpha = M - t_{\alpha/2} + 1 \Rightarrow$

$$\Pr\left(T^+ < C_\alpha\right) = \Pr\left(T^+ \le M - t_{\alpha/2}\right) = \Pr\left(T^+ \ge t_{\alpha/2}\right) = \frac{\alpha}{2} \approx \Pr\left(T^* < -z_{\alpha/2}\right)$$

$$= \Pr\left(\frac{T^+ - \mathrm{E}_0[T^+]}{\sqrt{\mathrm{Var}_0(T^+)}} < -z_{\alpha/2}\right) = \Pr\left(T^+ < \mathrm{E}_0[T^+] - z_{\alpha/2}\sqrt{\mathrm{Var}_0(T^+)}\right)$$

C_α can be approximated by

$$C_\alpha \approx \mathrm{E}_0[T^+] - z_{\alpha/2}\sqrt{\mathrm{Var}_0(T^+)} = \frac{n(n+1)}{4} - z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{2.18}$$

Example 2.6 Table 3.1 in Example 3.1 of the textbook (page 43) presents a set of paired data on Hamilton depression factor IV (the "suicidal" factor). The data X_i and Y_i, $i=1,\ldots,9$, represent the values pre and post a treatment on 9 patients.

The differences $Z_i = Y_i - X_i$, ranks R_i of $|Z_i|$ and indicators ψ_i of positive Z_i are calculated following the data on page 44, showing that $T^+ = r_1 + r_2 = 2+3=5$. The outcomes for $T^+ = 0,1,2,3,4,5$ are listed below.

T^+	0	1	2	3	4	5
Outcome	$B=0$	(1)	(2)	(3), (1,2)	(4), (1,3)	(5), (1,4), (2,3)

where (1) stands for $r_1 = 1$; (1,2) for $(r_1, r_2) = (1,2)$; and so on.

To test $H_0 : \theta = 0$ against $H_1 : \theta < 0$ (lower value means improvement),

$$p\text{-value} = \Pr(T^+ \le 5) = \sum_{k=0}^{5} \Pr(T^+ = k) = \frac{1+1+1+2+2+3}{2^9} = \frac{10}{2^9} = 0.0195$$

Hence H_0 is rejected at the 5% level in favour of $\theta < 0$. This provides sufficient evidence that the treatment is effective.

The approximate rejection rule at the 5% level is $T^* < 1.645$. Since

$$T^* = \frac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = \frac{5 - 9\times 10/4}{\sqrt{9\times 10\times 19/24}} = -2.073 < -1.645 = -z_{0.05},$$

the approximate rule rejects H_0 at the 5% level and concludes $\theta < 0$ as well.

Since $M = 9\times 10/2 = 45$, θ is estimated by $\hat{\theta} = W_{((M+1)/2)} = W_{(23)}$ in (2.15).

To obtain a confidence interval of θ at a desired level 95%, first calculate using $\Pr(T^+ \ge t) = \Pr(T^+ \le M - t)$:

$$\Pr(T^+ \ge 40) = \Pr(T^+ \le 45 - 40) = \Pr(T^+ \le 5) = 0.0195 < 0.025 \;\Rightarrow\; t_{0.0195} = 40$$

Since $T^+ = 6$ for $r_1 = 6$, $(r_1, r_2) = (1,5), (2,4)$ and $(r_1, r_2, r_3) = (1,2,3)$,

$$\Pr(T^+ \ge 39) = \Pr(T^+ \le 6) = \Pr(T^+ \le 5) + \Pr(T^+ = 6) = \frac{10+4}{2^9} = 0.0273 > 0.025$$

Thus the smallest achievable level over 95% is $1 - 2(0.0195) = 96.1\%$ ($\alpha = 0.039$).

Calculate $C_{0.039} = M + 1 - t_{0.0195} = 45 + 1 - 40 = 6$. By (2.17), a 96.1% confidence interval for θ is given by $(\theta_L, \theta_U) = \left(W_{(C_{0.039})}, W_{(t_{0.0195})}\right) = \left(W_{(6)}, W_{(40)}\right)$.

The ordered values $W_{(1)}, \ldots, W_{(45)}$ of Walsh averages defined in (2.14) are listed in the next page, from which we get

$$\hat{\theta} = W_{(23)} = -0.4600 \quad \text{and} \quad (\theta_L, \theta_U) = \left(W_{(6)}, W_{(40)}\right) = (-0.7860, -0.0100)$$

If we use the approximation in (2.18), then

$$C_{0.05} \approx \frac{9 \times 10}{4} - 1.96\sqrt{\frac{9 \times 10 \times 19}{24}} = 5.96$$

This would round to the exact value $C_{0.05} = 6$ and lead to the same confidence interval $(-0.7860, -0.0100)$ as above.

More details on the Hamilton depression factor IV data in this example can be found in Examples 3.1, 3.3 and 3.4 of the textbook.

Ordered values of Walsh averages for paired data on Hamilton depression factor IV in Examples 3.1 and 3.3 of the textbook

Order (i)	$W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(45)}$				
(1) – (5)	−1.0220	−0.9870	−0.9520	−0.8210	−0.8060
(6) – (10)	−0.7860	−0.7710	−0.7560	−0.7260	−0.7210
(11) – (15)	−0.6910	−0.6200	−0.6050	−0.5900	−0.5550
(16) – (20)	−0.5400	−0.5250	−0.5160	−0.5100	−0.4900
(21) – (25)	−0.4810	−0.4710	−0.4600	−0.4375	−0.4360
(26) – (30)	−0.4300	−0.4025	−0.3150	−0.3000	−0.2700
(31) – (35)	−0.2550	−0.2500	−0.2365	−0.2215	−0.2200
(36) – (40)	−0.2050	−0.1750	−0.1715	−0.1415	−0.0100
(41) – (45)	0.0350	0.0685	0.0800	0.1135	0.1470