One-sample and paired tests

1 Distribution-Free vs. Nonparametric Methods

Nonparametric methods. A statistical method is called *nonparametric* if it does not assume that the underlying population distribution belongs to a finite-dimensional parametric family (such as the normal family indexed by (μ, σ^2)). Instead, the distribution is treated as an unknown, possibly infinite-dimensional object. Nonparametric methods may still rely on structural assumptions such as independence, continuity, symmetry, or smoothness.

Examples: rank-based tests, kernel density estimation, nonparametric regression, spline smoothing, and the empirical distribution function.

Distribution-free methods. In classical usage, distribution-free methods are almost always statistical tests. A statistical test is called *distribution-free* if the null distribution of its test statistic does not depend on the underlying population distribution (within a specified class, such as continuous distributions). In this case, critical values and p-values are valid without knowing the exact form of the population distribution.

Examples: the sign test, Wilcoxon signed-rank test, Mann–Whitney U test, and the Kolmogorov–Smirnov test (under continuity).

Key distinction. Nonparametric refers to the *model* (no parametric family is assumed), whereas distribution-free refers to the *sampling distribution of the test statistic*. All classical distribution-free tests are nonparametric, but many nonparametric methods are not distribution-free.

1.1 Asymptotic Distribution-Free Tests

Many commonly used statistical tests are not distribution-free in finite samples, but become distribution-free *asymptotically*. Specifically, a test statistic T_n is called *asymptotically distribution-free* if

$$T_n \xrightarrow{d} T,$$

where the limiting distribution T does not depend on the underlying population distribution.

This phenomenon is typically a consequence of limit theorems such as the Central Limit Theorem, functional central limit theorems, or likelihood-based asymptotic results.

Examples:

- Wald or Z tests, where the standardized estimator converges to $N(0, 1)$.
- Likelihood ratio tests, which converge to a χ^2 distribution under regularity conditions (Wilks' theorem).
- Many rank- and permutation-based tests, whose limiting null distributions are universal.

Remark. Asymptotic distribution-freeness holds only under appropriate regularity conditions (e.g., independence, finite variance, smoothness). In finite samples, these tests generally depend on the underlying data-generating distribution.

Summary:

- Nonparametric methods avoid specifying a parametric family for the population distribution.
- Distribution-free methods have test statistics whose null distributions do not depend on the population distribution.
- Many parametric and nonparametric tests are distribution-free only in the asymptotic sense.
- The distribution-free property is a key aspect of 'many' nonparametric procedures.

Distribution-free asymptotic test statistic

- Test statistic: $T(\cdot) = T(X_1, \ldots, X_n)$, a function of the data.

Example

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}, \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1},$$

where μ is known under H_0.

Distribution-free test statistic

- $T_1 = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
- $T_2 = \dfrac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

The distributions of T_1 and T_2 hold for any distributional properties of the data.

Nonparametric distribution-free test statistic

- The class U, $T(\cdot)$ is distribution free over contains more than one distributional forms.
- Distribution-free confidence interval, distribution-free multiple comparison procedure, distribution-free confidence band, asymptotically distribution-free test statistic, asymptotically distribution-free multiple comparison procedure, and asymptotically distribution-free confidence band.

2 Rank statistic

- **Absolute rank:** For any random variable $Z_1, \ldots, Z_n$, the absolute rank of Z_i, denoted by R_i, is the rank of $|Z_i|$ among $|Z_1|, \ldots, |Z_n|$.
- **Rank statistic:** A statistic $T(R)$ based only on the ranks of a sample is a rank statistic.
  - $T(R)$ is distribution-free over i.i.d. joint continuous distribution.

- **Signed rank:** The signed rank of Z_i is $R_i\psi_i$, where

$$\psi_i = \begin{cases} 1, & Z_i > 0, \\ 0, & Z_i < 0. \end{cases}$$

- **Signed rank statistic:** A statistic $T(\psi, R) = T(R_1\psi_1, \ldots, R_n\psi_n)$ that is a function of $Z_1, \ldots, Z_n$ only through the signed ranks is a signed rank statistic.
  - $T(\psi, R)$ is distribution-free over i.i.d. joint continuous distribution symmetric about 0.

3 Sign test (Fisher) – paired replicates data / one-sample data

Sign test

- $Z_1, \ldots, Z_n$ random sample from a continuous population that has a common median θ.
- If $Z_i \sim F_i$, then $F_i(\theta) = P(Z_i \le \theta) = P(Z_i > \theta) = 1 - F_i(\theta)$.
- Hypothesis testing:

$$H_0 : \theta = 0 \qquad \text{versus} \qquad H_A : \theta \neq 0.$$

Sign test (Cont.)

- Sign test statistic:

$$B = \sum_{i=1}^{n} \psi_i.$$

- Motivation:
  - When θ is larger than 0, there will be larger number of positive Z_i's $\rightarrow$ big B value $\rightarrow$ reject H_0 in favor of $\theta > 0$.
- Under H_0, $B \sim \text{Binomial}(n, 1/2)$.
- Significance level α: probability of rejecting H_0 when it is true.
- Note:
  - choices of α are limited to possible values of the $B \sim \text{Binomial}(n, 1/2)$ cdf
  - compare the distribution of B under H_0 and the observed test statistic value.

Rejection regions

- $H_A : \theta > 0$, Reject H_0 if $B \ge b_{\alpha;n,1/2}$.
- $H_A : \theta < 0$, Reject H_0 if $B \le n - b_{\alpha;n,1/2}$.
- $H_A : \theta \neq 0$, Reject H_0 if $B \ge b_{\alpha/2;n,1/2}$ or $B \le n - b_{\alpha/2;n,1/2}$.

Large-Sample Approximation (Sign test)

$$B^* = \frac{B - E_0(B)}{V_0(B)^{1/2}} \sim N(0,1) \quad \text{as } n \to \infty,$$

where

$$E_0(B) = \frac{n}{2}, \qquad V_0(B) = \frac{n}{4}.$$

Rejection regions:

- $H_A : \theta > 0$, Reject H_0 if $B^* \geq z_\alpha$.
- $H_A : \theta < 0$, Reject H_0 if $B^* \leq -z_\alpha$.
- $H_A : \theta \neq 0$, Reject H_0 if $B^* \geq z_{\alpha/2}$ or $B^* \leq -z_{\alpha/2}$.

Ties (Sign test)

- Discard zero Z values and redefine n.
- If too many zeros, choose alternative statistical procedure (Chapter 10).

4 Example (Sign test)

Example (HWC: Chapter 3, Example 3.5, pg. 65) – paired sample sign test

- Beak-Clapping Counts.
- Subjects: chick embryos.
- X = average number of claps per minute during the dark period.
- Y = average number of claps per minute during the period of illumination.
- Test responsivity of a chick embryo to a light stimulus.
- $H_A : \theta > 0$.

```
df = data.frame(X = c (5.8,13.5,26.1,7.4,7.6,23,10.7,9.1,
19.3,26.3,17.5,17.9,18.3,14.2,55.2,15.4,30,21.3,
26.8 , 8.1 , 24.3 , 21.3,18.2,22.5,31.1),
Y = c(5,21,73,25,3,77,59,13,36,46,9,25,
59,38,70,36,55,46,25,30,29,46,71,31,33))
head(df)
```

```
library(dplyr)
df = mutate(df, Z= Y-X, Psi = ifelse(Z > 0 , 1 , 0 ) )
head(df)
```

- lower.tail=F provides $P(B > b_{\alpha=.05}) = .05$

```
qbinom(p = .05, size = length(df$Psi),
      prob = 1/2, lower.tail = FALSE)
```

- We need $P(B \geq b) = .05$. Therefore, Reject H_0 if $B \geq 18$. However, the significance level is not .05.

```
1 - pbinom((18-1), size = length(df$Psi),
          prob = 1/2, lower.tail = TRUE)
```

- Observed value of test statistic is

```
sum(df$Psi)
```

- We reject in favor of $\theta > 0$ at the $\alpha = .05$ level.
- Didn't use actual Z_i.
- Actual magnitude of the Z_i's will be necessary for distribution-free point and interval estimates of θ associated with the sign test.

Built-in function `SIGN.test` in package `BSDA`

```
library(BSDA)
SIGN.test(df$Y, df$X, alt = "greater")
```

P-value using `pbinom` and large-sample approximation

```
1 - pbinom((21-1), size = length(df$Psi),
          prob = 1/2, lower.tail = TRUE)
B.star <- (21-25/2)/sqrt(25/4)
B.star
1-pnorm(B.star)
```

- Both the exact test and the large-sample approximation indicate that there is strong evidence that chick embryos are indeed responsive to a light stimulus, as measured by an increase in the frequency of beak-claps.
- To test $H_0 = \theta_0$, compute $Z_1 - \theta_0, \ldots, Z_n - \theta_0$ and do sign test on the Z's.

5 Parametric t-test

Q: Is rank test always less useful then parametric t-test?

- Let $Z_i \sim N(\theta, \sigma^2)$.
- $H_0 : \theta = 0$ versus $H_A : \theta > 0$.
- Test statistic:
$$T = \frac{\bar{Z} - \theta}{s/\sqrt{n}}.$$
- T is Studentized t-distribution with degrees of freedom $n - 1$.
- t_0: the observed value of test statistic.
- P-value: $P(T \geq t_0)$.

6 Wilcoxon signed rank test

Assumptions

$$Z_i = Y_i - X_i \sim F, \quad \text{where } F \text{ is symmetric about common median } \theta.$$

Test statistic

Let

$$T^+ = \sum_{i=1}^{n} R_i \psi_i,$$

sum of positive signed ranks.

- no-closed form distribution.
- use iterative algorithms.

We have $T^+ + T^- = \frac{n(n+1)}{2}$. Under null, we must have $T^+ = T^-$ and thus if null is true, we should have $T^+ = T^- \approx \frac{n(n+1)}{4}$. This forms the basis for the following rejection regions.

Rejection regions

- $H_A : \theta > 0$, Reject H_0 if $T^+ \geq t_\alpha$.
- $H_A : \theta < 0$, Reject H_0 if $T^+ \leq \dfrac{n(n+1)}{2} - t_\alpha$.
- $H_A : \theta \neq 0$, Reject H_0 if $T^+ \geq t_{\alpha/2}$ or $T^+ \leq \dfrac{n(n+1)}{2} - t_{\alpha/2}$.

Ties

- where the constant t_α is chosen to make the type I error probability equal to α. The tests can be performed using the R command `wilcox.test`. The t_α critical values can be obtained from the R command `psignrank`.
- Discard zero values among the Z_i's.
- If there are ties, assign each of the observations in a tied group the average of the integer ranks that are associated with the tied group.
- not exact test

7 Exact Wilcoxon Signed-Rank p-Values

We describe how exact p-values for the Wilcoxon signed-rank test are computed.

7.1 Null distribution of the statistic

Let $Z_1, \dots, Z_n$ be paired differences from a continuous distribution symmetric about zero. Let $r_i = \text{rank}(|Z_i|)$ and define

$$T^+ = \sum_{i=1}^{n} r_i \mathbf{1}\{Z_i > 0\}.$$

Under the null hypothesis:

- the ranks are $\{1, 2, \dots, n\}$ with probability one,
- the signs are i.i.d. Bernoulli(1/2),
- each of the 2^n sign configurations is equally likely.

Hence, T^+ has the same distribution as the sum of a random subset of $\{1, 2, \dots, n\}$.

7.2 Exact probability mass function

Let S be a subset of $\{1, \dots, n\}$ such that $Z_i > 0$ if $i \in S$ and define

$$N_n(t) = \#\left\{ S \subset \{1, \dots, n\} : \sum_{i \in S} i = t \right\}.$$

Then the exact null distribution is

$$\Pr(T^+ = t) = \frac{N_n(t)}{2^n}, \quad t = 0, 1, \dots, \frac{n(n+1)}{2}.$$

7.3 Exact p-values

Let $t_{\max} = n(n+1)/2$ and let t_{obs} be the observed value of T^+.

- **Left-sided test:**
$$p = \Pr(T^+ \le t_{\text{obs}}).$$
- **Right-sided test:**
$$p = \Pr(T^+ \ge t_{\text{obs}}).$$
- **Two-sided test:**
$$p = 2 \min\left\{ \Pr(T^+ \le t_{\text{obs}}),\ \Pr(T^+ \ge t_{\max} - t_{\text{obs}}) \right\}.$$

7.4 Computation

Exact probabilities can be obtained by enumeration for small n or by dynamic programming for moderate n. In `R`, these calculations are implemented in the function `psignrank`, which evaluates the cumulative distribution function of T^+ under the null hypothesis.

The Wilcoxon signed-rank test is distribution-free because its null distribution depends only on n and not on the underlying data distribution.

8 Theoretical distribution of T^+

Large-Sample Approximation of the Wilcoxon Signed-Rank Test

Setup

Let

$$Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} F$$

be paired differences (or one-sample centered observations). We consider the hypothesis

$$H_0: \ \text{median}(Z) = 0 \quad \text{versus} \quad H_1: \ \text{median}(Z) \neq 0.$$

Assume throughout that under H_0, the distribution of Z_i is continuous and symmetric about zero, so that

$$\mathbb{P}(Z_i = 0) = 0 \quad \text{and ties occur with probability zero.}$$

Role of symmetry. The Wilcoxon signed-rank test conditions on the absolute values $|Z_1|, \ldots, |Z_n|$ and their ranks. Under the null hypothesis that the distribution of Z_i is symmetric about 0,

$$\Pr(Z_i > 0 \mid |Z_i|) = \Pr(Z_i < 0 \mid |Z_i|) = \tfrac{1}{2},$$

so the signs are i.i.d. Bernoulli(1/2) and independent of the ranks. This symmetry assumption is what makes the null distribution of the signed-rank statistic distribution-free.

Define

$$A_i = |Z_i|, \qquad S_i = \text{sign}(Z_i) \in \{-1, +1\},$$

and let R_i denote the rank of A_i among $A_1, \ldots, A_n$, so that

$$R_i \in \{1, \ldots, n\}, \qquad \sum_{i=1}^{n} R_i = \frac{n(n+1)}{2}.$$

The Wilcoxon signed-rank statistic is defined as

$$T = \sum_{i: Z_i > 0} R_i = \sum_{i=1}^{n} R_i \mathbf{1}\{S_i = +1\}.$$

Key Distributional Property Under H_0

Under H_0, conditional on the absolute values $(A_1, \ldots, A_n)$ (and hence on the ranks $(R_1, \ldots, R_n)$), the signs $S_1, \ldots, S_n$ are independent Rademacher random variables:

$$\mathbb{P}(S_i = +1 \mid A_1, \ldots, A_n) = \mathbb{P}(S_i = -1 \mid A_1, \ldots, A_n) = \frac{1}{2}.$$

Equivalently,

$$X_i := \mathbf{1}\{S_i = +1\} \quad \text{are i.i.d. Bernoulli} \left(\frac{1}{2}\right),$$

independent of the ranks.

Mean of the Signed-Rank Statistic

Conditioning on the ranks,

$$\mathbb{E}[T \mid R_1, \ldots, R_n] = \sum_{i=1}^{n} R_i \mathbb{E}[X_i] = \frac{1}{2} \sum_{i=1}^{n} R_i = \frac{n(n+1)}{4}.$$

Since this quantity is nonrandom,

$$\boxed{\mathbb{E}_0[T] = \frac{n(n+1)}{4}.}$$

Variance of the Signed-Rank Statistic

Again conditioning on the ranks and using independence,

$$\operatorname{Var}(T \mid R_1, \ldots, R_n) = \sum_{i=1}^{n} R_i^2 \operatorname{Var}(X_i) = \frac{1}{4} \sum_{i=1}^{n} R_i^2.$$

Using the identity

$$\sum_{r=1}^{n} r^2 = \frac{n(n+1)(2n+1)}{6},$$

we obtain

$$\boxed{\operatorname{Var}_0(T) = \frac{n(n+1)(2n+1)}{24}.}$$

Asymptotic Normal Approximation

Write the centered statistic as

$$T - \mathbb{E}[T \mid R_1, \ldots, R_n] = \sum_{i=1}^{n} R_i \left(X_i - \frac{1}{2} \right).$$

This is a sum of independent, mean-zero random variables with total variance

$$s_n^2 = \frac{n(n+1)(2n+1)}{24}.$$

Since $\max_i R_i = n$ and $s_n^2 = O(n^3)$, the Lindeberg condition holds, and by the Lindeberg–Feller central limit theorem,

$$\frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Thus, for large n,

$$\boxed{T \approx \mathcal{N}\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24} \right) \quad \text{under } H_0.}$$

Continuity Correction

Because T is discrete, a continuity-corrected test statistic is often used:

$$Z_{\rm cc} = \frac{T - \frac{n(n+1)}{4} \mp 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24}}},$$

where the sign is chosen to move the numerator toward zero.

Alternative Signed Form

Define the signed-rank sum

$$W = \sum_{i=1}^{n} R_i S_i.$$

Since $S_i = 2X_i - 1$,

$$W = 2T - \frac{n(n+1)}{2}.$$

Under H_0,

$$\mathbb{E}[W] = 0, \qquad \mathrm{Var}(W) = \frac{n(n+1)(2n+1)}{6},$$

and

$$\frac{W}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Remarks on Zeros and Ties

If some $Z_i = 0$, they are removed and the above results apply with n replaced by $n' = \#\{i : Z_i \neq 0\}$. If ties occur among $|Z_i|$, average ranks are used and the variance is reduced accordingly.

9 Exact Distribution of the Wilcoxon Signed-Rank Statistic for $n = 3$

We illustrate how the exact null distribution of the Wilcoxon signed-rank statistic can be derived by enumeration and verified via Monte Carlo simulation.

9.1 Definition of the statistic

Let $Z_1, \dots, Z_n$ be i.i.d. observations from a continuous distribution symmetric about zero. Let $r_i = \mathrm{rank}(|Z_i|)$ denote the rank of $|Z_i|$ among $\{|Z_1|, \dots, |Z_n|\}$. The Wilcoxon signed-rank statistic is

$$T^+ = \sum_{i=1}^{n} r_i \, \mathbf{1}\{Z_i > 0\}.$$

For $n = 3$ and no ties, the ranks are $\{1, 2, 3\}$.

9.2 Enumeration of all possible outcomes

Under the null hypothesis of symmetry:

- the signs $\mathbf{1}\{Z_i > 0\}$ are i.i.d. Bernoulli(1/2),
- the signs are independent of the absolute values and hence of the ranks.

Therefore, conditional on the ranks $\{1, 2, 3\}$, there are $2^3 = 8$ equally likely sign configurations. We enumerate all sign patterns $(b_1, b_2, b_3) \in \{0, 1\}^3$, where $b_i = \mathbf{1}\{Z_i > 0\}$, and compute

$$T^+ = 1 \cdot b_1 + 2 \cdot b_2 + 3 \cdot b_3.$$

```
library(gtools)

x <- c(0,1)
df <- permutations(n = 2, r = 3, v = x,
                   repeats.allowed = TRUE) %>% data.frame
df
```

We then compute the corresponding values of T^+:

```
T.plus <- apply(df, 1,
                function(x) sum(x * seq(1,3)))
df <- mutate(df, T.plus = T.plus)
df
```

Finally, we tabulate the exact probability mass function:

```
table(df$T.plus) / sum(table(df$T.plus))
```

9.3 Exact null distribution

The statistic T^+ can take values

$$T^+ \in \{0, 1, 2, 3, 4, 5, 6\}.$$

Each value corresponds to a subset sum of $\{1, 2, 3\}$. Counting the number of subsets that produce each sum yields:

$$\Pr(T^+ = t) = \begin{cases} 1/8, & t \in \{0, 1, 2, 4, 5, 6\}, \\ 1/4, & t = 3. \end{cases}$$

The larger probability at $t = 3$ arises because two distinct sign patterns, $\{3\}$ and $\{1, 2\}$, yield the same value.

9.4 Monte Carlo verification

We now verify the theoretical distribution via simulation. We generate i.i.d. $N(0, 1)$ samples of size $n = 3$, compute the signed-rank statistic for each sample, and estimate its empirical distribution.

```
n <- 3
nsim <- 10000

Z <- matrix(rnorm(n * nsim), ncol = n)
```

```
T.plus.mc <- apply(Z, 1, function(x) {
  sum(rank(abs(x)) * (x > 0))
})

table(T.plus.mc) / nsim
```

9.5 Interpretation

Because the normal distribution is continuous and symmetric:

- ties occur with probability zero,
- ranks are a permutation of $\{1, 2, 3\}$,
- signs are independent of the ranks.

Thus, the Monte Carlo distribution converges to the exact distribution obtained by enumeration, confirming the validity of the signed-rank null distribution for $n = 3$.

This enumeration argument generalizes to arbitrary n, where the exact null distribution of T^+ is the distribution of subset sums of $\{1, \dots, n\}$, with each subset occurring with probability 2^{-n}.

10 Example: Wilcoxon Signed-Rank Test

We illustrate the Wilcoxon signed-rank test using data from a paired clinical study.

10.1 Study description

- Data are collected from $n = 9$ patients who received a tranquilizer.
- X (pre) denotes the factor IV value measured at the first visit, shortly after initiation of therapy.
- Y (post) denotes the factor IV value measured at the second visit.
- Clinical improvement corresponds to a *reduction* in factor IV values.

The scientific question is whether the tranquilizer leads to improvement, i.e., whether post-treatment values tend to be smaller than pre-treatment values.

10.2 Formulation of hypotheses

Let $Z_i = Y_i - X_i$ denote the paired differences. The hypotheses are:

$$H_0 : \text{the distribution of } Z_i \text{ is symmetric about } 0,$$

$$H_1 : Z_i < 0 \quad \text{(systematic reduction after treatment)}.$$

This corresponds to a one-sided Wilcoxon signed-rank test.

10.3 Wilcoxon signed-rank test in R

```
pre = c(1.83, .50, 1.62, 2.48, 1.68, 1.88,
       1.55, 3.06, 1.30)

post = c(.878, .647, .598, 2.05, 1.06, 1.29,
        1.06, 3.14, 1.29)

wilcox.test(post, pre, paired = TRUE,
           alternative = "less")
```

Here:

- `paired = TRUE` specifies a paired design,
- `alternative = "less"` corresponds to testing $Y < X$, i.e., post-treatment values are smaller.

10.4 Manual construction of the test statistic

To understand the test, we explicitly compute the signed-rank statistic.

```
df <- data.frame(X = pre, Y = post)

df <- mutate(df,
            Z = Y - X,
            R = rank(abs(Z)),
            psi = ifelse(Z > 0, 1, 0),
            Rpsi = R * psi)

df
```

For each patient:

- $Z = Y - X$ is the paired difference,
- R is the rank of $|Z|$ among all absolute differences,
- $\psi = \mathbf{1}\{Z > 0\}$ indicates a positive difference,
- $R\psi$ contributes to the signed-rank statistic.

The Wilcoxon signed-rank statistic is

$$T^+ = \sum_{i=1}^{n} R_i \mathbf{1}\{Z_i > 0\} = \sum_{i=1}^{n} R_i \psi_i.$$

10.5 Exact p-value computation

Under the null hypothesis of symmetry, the exact distribution of T^+ depends only on n and not on the underlying distribution.

The observed value is $T^+ = \sum_i R_i \psi_i = 5$. The exact one-sided p-value is

$$p = \Pr(T^+ \leq 5).$$

```
psignrank(q = sum(df$Rpsi), n = 9, lower.tail = TRUE)
```

Here, `psignrank` evaluates the cumulative distribution function of the Wilcoxon signed-rank statistic under the null hypothesis.

10.6 Conclusion

- The resulting p-value is less than $\alpha = 0.05$.
- We reject the null hypothesis of no treatment effect.
- There is strong evidence that the tranquilizer leads to patient improvement, as measured by a reduction in Hamilton scale factor IV values.

This example illustrates how the Wilcoxon signed-rank test provides *exact, distribution-free inference* for paired data under symmetry, while remaining robust to outliers and non-normality.

11 Point and Interval Estimates

In addition to hypothesis testing, classical and nonparametric procedures provide associated point estimates and confidence intervals for the location parameter θ.

For paired data or one-sample location problems, the three commonly used procedures—the sign test, the Wilcoxon signed-rank test, and the t-test— each lead to a natural estimator and confidence interval for θ.

11.1 Preliminaries: order statistics and quantiles

Let $Z_1, \ldots, Z_n$ be a sample from a continuous distribution.

- The *order statistics* are
$$Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}.$$
- $Z_{(1)}$ is the sample minimum and $Z_{(n)}$ is the sample maximum.
- A *quantile* is a value that splits the distribution into regions with equal probability mass.

Order statistics play a central role in nonparametric estimation and confidence interval construction.

11.2 Point and interval estimation associated with the sign test

The sign test is a test for the median θ of a symmetric distribution. Its associated point estimator is the sample median.

11.2.1 Point estimate

The point estimate of θ is
$$\tilde{\theta} = \text{median}\{Z_1, \ldots, Z_n\}.$$

In terms of order statistics:

- If n is odd, then
$$\tilde{\theta} = Z_{(k+1)}, \quad k = \frac{n-1}{2}.$$
- If n is even, then
$$\tilde{\theta} = \frac{Z_{(k)} + Z_{(k+1)}}{2}, \quad k = \frac{n}{2}.$$

11.2.2 Confidence interval

A $100(1-\alpha)\%$ confidence interval for θ associated with the two-sided sign test is

$$\left(Z_{(n+1-b_{\alpha/2;n,1/2})},\ Z_{(b_{\alpha/2;n,1/2})}\right),$$

where $b_{\alpha/2;n,1/2}$ is the upper $\alpha/2$ percentile of the Binomial$(n, 1/2)$ distribution (the null distribution of the sign test statistic).

This confidence interval is distribution-free and depends only on the order statistics, not on any parametric assumptions.

11.3 Point and interval estimation associated with the Wilcoxon signed-rank test

The Wilcoxon signed-rank test is associated with the *Hodges–Lehmann estimator*, a robust estimator of the location parameter.

11.3.1 Point estimate: Hodges–Lehmann estimator

The Hodges–Lehmann estimator of θ is defined as

$$\hat{\theta} = \text{median}\left\{\frac{Z_i + Z_j}{2} : 1 \le i \le j \le n\right\}.$$

The quantities

$$\frac{Z_i + Z_j}{2}, \quad i \le j,$$

are called *Walsh averages.*

- There are $M = n(n+1)/2$ Walsh averages.
- Let
$$W_{(1)} \le W_{(2)} \le \cdots \le W_{(M)}$$
denote the ordered Walsh averages.

The Hodges–Lehmann estimator is then:

- If M is odd,
$$\hat{\theta} = W_{(k+1)}, \quad k = \frac{M-1}{2}.$$
- If M is even,
$$\hat{\theta} = \frac{W_{(k)} + W_{(k+1)}}{2}, \quad k = \frac{M}{2}.$$

11.3.2 Confidence interval

A $100(1-\alpha)\%$ confidence interval for θ associated with the two-sided Wilcoxon signed-rank test is

$$\left(W_{\left(\frac{n(n+1)}{2}+1-t_{\alpha/2}\right)},\ W_{(t_{\alpha/2})}\right),$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ percentile of the null distribution of the Wilcoxon signed-rank statistic T^+.

The percentile points $t_{\alpha/2}$ can be obtained using the `R` function `psignrank`.

11.4 Comparison and interpretation

- The sign-test interval targets the median and is extremely robust, but may be wide.
- The Wilcoxon interval targets a location parameter under symmetry and is typically shorter and more efficient.
- Both intervals are exact and distribution-free under their respective assumptions.

The Hodges–Lehmann estimator can be viewed as a compromise between the sample mean and the sample median, offering high efficiency under symmetry while retaining robustness to outliers.

12 Relationship between Wilcoxon signed rank test statistic and Walsh averages (Tukey (1949))

- HWC page 57, comment 17.
- Wilcoxon test statistic: $T^+ = \sum_{i=1}^n R_i \psi_i$.
- Number of Walsh averages greater than $\hat{\theta}$:

$$W^+ = \# \left\{ \frac{Z_i + Z_j}{2} > \hat{\theta} \right\}.$$

- Prove $T^+ = W^+$ by induction.
- Base of the Induction:
  - Assume that $\hat{\theta}$ is greater than all $Z_1, \ldots, Z_n$, then $\hat{\theta}$ is greater than all Walsh averages. Thus, $W^+ = 0$.
  - Then, $Z_i - \hat{\theta}$ are all negative. Thus, $T^+ = 0$.

Induction Steps

- Move $\hat{\theta}$ to the left passing through $Z_1, \ldots, Z_n$ one and two at the time and show that
  - W^+ changes value when moving past a Walsh average by the same amount.
  - T^+ changes value when
    - ranks of some $|Z_i - \hat{\theta}|$ change or
    - sign of some rank change by the same amount.

13 Comparison

- Power of a statistical test: the probability of rejecting the null hypothesis when it is false.
- The power of the sign test can be low relative to t-test.
- The power of signed-rank Wilcoxon test is nearly that of the t-test for normal distributions and generally greater than that of the t-test for distributions with heavier tails than the normal distribution.

Note: Read HWC page 71, comment 35 (power results for sign test).

14 Empirical power calculation using asymptotic distribution

```
power.compute <- function(n = 30,
df = 2,
nsims = 1000,
theta = 0){
wil.sign.rank = rep(0, nsims)
ttest = rep(0,nsims)
Z = matrix((rt(n*nsims,df) + theta),
ncol = n,nrow = nsims)
wil.sign.rank = apply(Z, 1 , function(x){
wilcox.test(x)$p.value})
ttest = apply(Z, 1 , function(x){t.test(x)$p.value})
pow.wil.sign.rank = mean(wil.sign.rank <=.05)
pow.ttest = mean(ttest <=.05)
rt = c(pow.wil.sign.rank, pow.ttest)
names(rt) = c("Wilcoxon.signed.rank.power",
"t.test.power")
return(rt)
}
```

$\theta = 0$

```
power.compute.val = power.compute(n=30, df =2,
nsims =1000, theta = 0)
power.compute.val
```

$\theta = 0.5$

```
power.compute.val = power.compute(n=30, df =2,
nsims =1000, theta = 0.5)
power.compute.val
```

$\theta = 1$

```
power.compute.val = power.compute(n=30, df =2,
nsims =1000, theta = 1)
power.compute.val
```

15 Summary

- Assumptions on F_i:
  - Sign Test: any continuous distribution.
  - Signed-Rank Test: any symmetric continuous distribution.
  - t-test: any normal distribution.

- The continuity assumption assures that ties are impossible: With probability one we have $Z_i \neq Z_j$ when $i \neq j$.
- The continuity assumption is only necessary for exact hypothesis tests not for estimates and confidence intervals.

16 Sign Test vs. Wilcoxon Signed-Rank Test

Both the sign test (often called the Fisher exact sign test) and the Wilcoxon signed-rank test are nonparametric procedures for analyzing paired data or one-sample location problems. They differ in how much information from the data they use.

16.1 Problem setup

Let $Z_1, \ldots, Z_n$ be paired differences (or one-sample observations after centering). We are interested in inference on a location parameter θ, typically interpreted as a median or center of symmetry.

16.2 Sign test (Fisher exact sign test)

Test statistic. The sign test uses only the signs of the differences:

$$B = \sum_{i=1}^{n} \mathbf{1}\{Z_i > 0\}.$$

Null hypothesis.

$$H_0: \ \Pr(Z_i > 0) = \Pr(Z_i < 0) = \tfrac{1}{2},$$

which is equivalent to θ being the median of the distribution.

Null distribution. Under H_0,

$$B \sim \text{Binomial}(n, 1/2),$$

exactly, for any sample size.

Key features.

- Uses only the direction (sign) of each difference.
- Requires no assumptions beyond continuity at zero.
- Exact finite-sample inference.
- Very robust but low power.

16.3 Wilcoxon signed-rank test

Test statistic. The Wilcoxon signed-rank test uses both sign and magnitude information:

$$T^+ = \sum_{i=1}^{n} r_i \mathbf{1}\{Z_i > 0\},$$

where $r_i = \text{rank}(|Z_i|)$.

Null hypothesis.

$$H_0: \text{ the distribution of } Z_i \text{ is symmetric about } \theta.$$

Null distribution. Under H_0 and no ties, T^+ has a distribution given by subset sums of $\{1, \ldots, n\}$. Exact p-values are available for small and moderate n.

Key features.

- Uses both sign and relative magnitude information.
- Assumes symmetry of the distribution.
- Exact finite-sample inference (with no ties).
- Substantially higher power than the sign test.

16.4 Comparison

	Sign Test	**Wilcoxon Signed-Rank**
Data used	Signs only	Signs + ranks of magnitudes
Target parameter	Median	Center of symmetry
Assumptions	Very minimal	Symmetry required
Exact for finite n	Yes	Yes (no ties)
Efficiency under normality	Low	High (ARE ≈ 0.955)
Robustness	Very high	High

16.5 Efficiency and power

Under normality, the asymptotic relative efficiency (ARE) of the Wilcoxon signed-rank test relative to the t-test is approximately 0.955, whereas the sign test has ARE approximately 0.637. Thus, the Wilcoxon test is far more powerful when symmetry holds.

16.6 When to use which

- Use the **sign test** when:
  - the sample size is very small,
  - extreme robustness is required,
  - symmetry cannot be reasonably assumed.
- Use the **Wilcoxon signed-rank test** when:
  - the distribution is plausibly symmetric,
  - you want much higher power than the sign test,
  - outliers may be present but not overwhelming.

The Wilcoxon signed-rank test can be viewed as an efficiency-improved version of the sign test that trades a mild symmetry assumption for substantially increased power.