One-sample or 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 $F_i(\theta) = P(Z_i \leq \theta) = P(Z_i > \theta) = 1 - F_i(\theta)$.
- 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, $B \sim (n, 1/2)$.
- Significance level α: probability of rejecting H_0 when it is true.
- Note:
  - choices of α are limited to possible values of the $B \sim (n, 1/2)$ 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 $B \geq b_{\alpha;n,1/2}$.
- $H_A : \theta < 0$, Reject H_0 if $B \leq n - b_{\alpha;n,1/2}$.
- $H_A : \theta \neq 0$, Reject H_0 if $B \geq b_{\alpha/2;n,1/2}$ or $B \leq n - b_{\alpha/2;n,1/2}$.

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 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 $S_i = \text{Sign}(Z_i)$

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

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 Theoretical distribution of T^+

Large-Sample Approximation of the Wilcoxon Signed-Rank Test

Setup

Let

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

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

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

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 $A_1, \dots, A_n$, so that

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

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 $(A_1, \dots, A_n)$ (and hence on the ranks $(R_1, \dots, R_n)$), the signs $S_1, \dots, S_n$ are independent Rademacher random variables:

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

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,

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

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:

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

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.

Enumerate all 2^n possible outcomes for sample size three $n = 3$:

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

```
T.plus = apply(df, 1,
function(x){sum(x %*% seq(1,3))} )
df = mutate(df, T.plus = T.plus)
df
```

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

8 Monte Carlo Simulation

Compare Monte Carlo simulation results with the theoretical results:

```
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 Example (Wilcoxon signed rank test)

- Data are from nine patients who received tranquilizer.
- X (pre) factor IV value was obtained at the first patient visit after initiation of therapy.
- Y (post) factor IV value was obtained at the second visit after initiation of therapy.
- Test improvement due to tranquilizer that corresponds to a reduction in factor IV values.

```
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,
           alterative = "less")
```

```
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
```

P-value is $P(T^+ \leq 5)$:

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

- There is strong evidence that tranquilizer does lead to patient improvement at $\alpha = .05$, as measured by a reduction in the Hamilton scale factor IV values.

10 Point and interval estimates

- All three tests (sign test, Wilcoxon signed rank, and t-test) have an associated estimate and confidence interval for the location parameter θ.
- Order statistic: $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$.
- $Z_{(1)}$ is the minimum.
- $Z_{(n)}$ is the maximum.
- Quantile: equally spaced splitting points of continuous intervals with equal probabilities.

Point and interval estimate of θ associated with the sign test statistic

- median: $\tilde{\theta} = \text{median}\{Z_i, i = 1, \ldots, n\}$.
- Let $Z_{(1)}, \ldots, Z_{(n)}$ denote the ordered Z_i and if
  - n is odd, $\tilde{\theta} = Z_{(k+1)}$, where $k = (n-1)/2$.
  - n is even, $\tilde{\theta} = \dfrac{Z_{(k)} + Z_{(k+1)}}{2}$, where $k = n/2$.

- $100(1-\alpha)\%$ confidence interval associated with two-sided test:

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

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

Point and interval estimate of θ associated with the Wilcoxon signed rank statistic

- Hodges–Lehmann estimator:

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

- Walsh averages: $(Z_i + Z_j)/2$, $i \le j$.
- $M = n(n+1)/2$ Walsh averages.
- $W_{(1)} \le \cdots \le W_{(M)}$ denote the ordered values of $(Z_i + Z_j)/2$.
- If
  - M is odd, $\hat{\theta} = W_{(k+1)}$, where $k = (M-1)/2$.
  - M is even, $\hat{\theta} = \dfrac{W_{(k)} + W_{(k+1)}}{2}$, where $k = M/2$.
- $100(1-\alpha)\%$ confidence interval associated with two-sided test:

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

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

- The percentile points can be found using the `R` function `psignrank`.

11 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.

12 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).

13 Empirical power calculation

```
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
```

14 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.