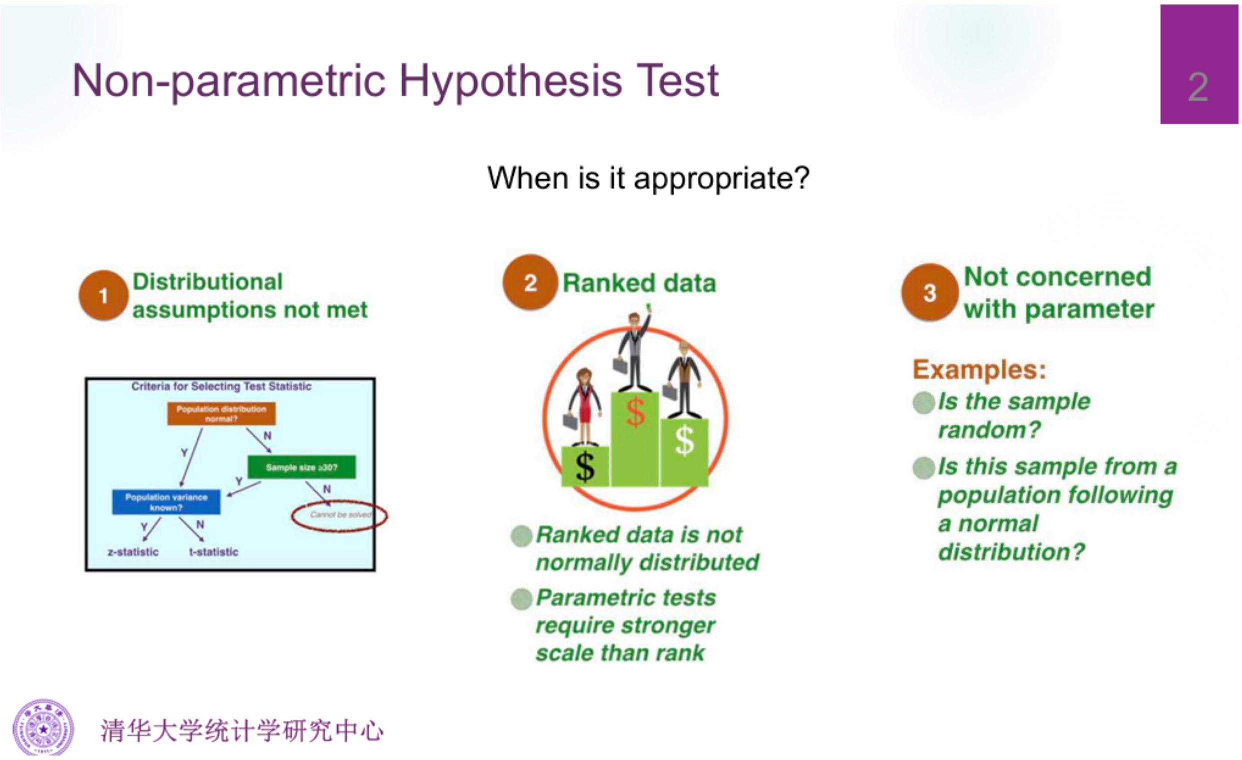
适用范围



Sign Test

1.1 Sign Test – Example

- Compare the effects of two soporific drugs

Subject	Drug 1	Drug 2	Diff (2-1)
1	1.9	0.7	−1.2
2	−1.6	0.8	2.4
3	−0.2	1.1	1.3
4	−1.2	0.1	1.3
5	−0.1	−0.1	0.0
6	3.4	4.4	1.0
7	3.7	5.5	1.8
8	0.8	1.6	0.8
9	0.0	4.6	4.6
10	2.0	3.4	1.4

Table: Hours of extra sleep on drugs 1 and 2, differences, signs and ranks of sleep study data

- Paired Comparison (成对比较); Paired data (成对数据)
- X_i, Y_i represent the hours of extra sleep on drugs 1 and 2 respectively
- $X_1, \cdots X_n$ independent; $Y_1, \cdots Y_n$ independent
- X_i, Y_i are dependent -- cannot use two sample t test
- Take the difference Diff(2-1): $Z_i = Y_i - X_i$ i.i.d.
- H_0: $\mu = 0 \leftrightarrow H_1$: $\mu \neq 0$ where $\mu = E(Z_i)$

1.2 Sign Test

- Sign test: analogue to the one sample t test 非参数版
- Used on paired data where the column of values represents differences (e.g., $Z_i = Y_i - X_i$)
- Sign test: also the simplest test for the median in the population

- t test不能用

1.1.1 Can we use t-test?

- T-tests: tests for the 【means】 of 【continuous】 data
  - One sample $H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$
  - Two sample $H_0: \mu_1 - \mu_2 = 0 \leftrightarrow H_1: \mu_1 - \mu_2 \neq 0$
- Underlying these tests is the assumption that the data arise from a normal distribution
- T-tests do not actually require normally distributed data to perform reasonably well in most circumstances
- But sometimes it goes wrong

-

Si

1.2 Sign Test

- Create a sign for each $Z_i, i = 1, \cdots, n$ $\quad S_i = \begin{cases} 1 & Z_i > 0 \\ -1 & Z_i < 0 \\ 0 & Z_i = 0 \end{cases}$
- H_0: The mean effects of two drugs are the same, or equivalently

Under H_0, what is the properties of S_i?

多选题 1分

设置

10

Create a sign for each $Z_i, i = 1, \cdots, n$

$$S_i = \begin{cases} 1 & Z_i > 0 \\ -1 & Z_i < 0 \\ 0 & Z_i = 0 \end{cases}$$

H_0: The mean effects of two drugs are the same, or equivalently.

Under H_0, what are the properties of S_i?

A The distribution of Z_i's is symmetric about its mean

B For any given subject, the difference in hours of sleep equally likely to be positive or negative, $P(S_i = 1) = P(S_i = -1) = 0.5$

C The number of positive Z_i's, n_+, can be regarded as binomial distribution.

D Let $n_0 = n_+ + n_-$, then $n_+ \sim B(n_0, \theta), H_0: \theta = 0.5 \leftrightarrow H_1: \theta \neq 0.5$

提交

星号部分-计算

1.2 Sign Test – technical details *

- The rejection region at level of significance α takes the form

$$D = \{\boldsymbol{X} = (X_1, \cdots, X_n) : n_+ \geq c \text{ or } n_+ \leq d\}$$

- The constants c, d can be determined by

$$\sum_{i=c}^{n_0} C_{n_0}^i 0.5^{n_0} \leq \frac{\alpha}{2}, \quad d = n_0 - c$$

- Compute the P-value of the test. Let $x_0 = \min\{n_+, n_0 - n^+\}$

$$p = \sum_{i=0}^{x_0} C_{n_0}^i 0.5^{n_0} + \sum_{i=n_0-x_0}^{n_0} C_{n_0}^i 0.5^{n_0}$$

(If n_0 is even and $n_+ = n_0/2$, then define $p = 1$.)

- Give a level of significance α, we reject H_0 if $p < \alpha$.

interpretation

- Drug example
  - $n_+ = 8, n_0 = 9$
  - (Exact) P-value (probability of observing 0,1,8,9 positives): $p = 0.0195 < 0.05$
  - Reject H_0 at level of significance $\alpha = 0.05$

R will help you

应用-Sign Test – test median

1.3 Sign Test – test median

- Sign test can be used to test about the median of a population

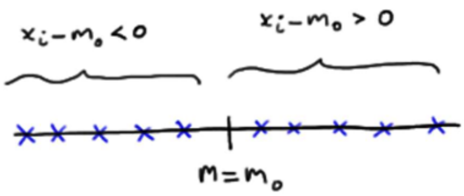
$$H_0: m = m_0$$

$$H_A: m > m_0 \text{ or } H_A: m < m_0 \text{ or } H_A: m \neq m_0$$

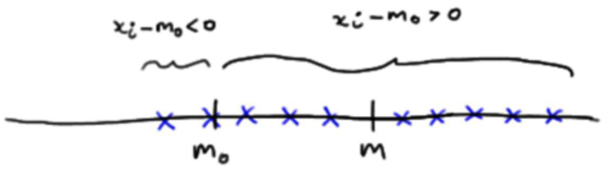
What is m ?

What is m_0?

If the **null hypothesis is true**, then we should expect about half of the $x_i - m_0$ quantities obtained to be positive and half to be negative:



If instead, $m > m_0$, then we should expect **more** than half of the $x_i - m_0$ quantities obtained to be positive and **fewer** than half to be negative:



1.3 Sign Test – test median

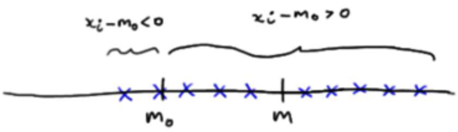
1. Calculate $X_i - m_0$ for $i = 1, 2, \ldots, n$.
2. Define $N-$ = the number of negative signs obtained upon calculating $X_i - m_0$ for $i = 1, 2, \ldots, n$.
3. Define $N+$ = the number of positive signs obtained upon calculating $X_i - m_0$ for $i = 1, 2, \ldots, n$.

If the null hypothesis is true, then $N-$ and $N+$ both follow a binomial distribution with parameters n and p = 1/2.

$$N- \sim b\left(n, \tfrac{1}{2}\right) \text{ and } N+ \sim b\left(n, \tfrac{1}{2}\right)$$

Suppose Ha: $m > m_0$,

we should reject the null hypothesis if n− is too small.



Or alternatively, if the P-value as defined by below is too small.

$$P = P(N- \leq n-)$$

Sign Test – Summary

- Sign test is appropriate if the population distribution (Z_i) is symmetric under H_0 or we are testing the about the median
- Similar arguments can be used to test hypothesis
  - H_0: $\theta \leq 0.5 \leftrightarrow H_1$: $\theta > 0.5$
- Drawback of sign test: it ignores magnitudes completely → it is inefficient (low power)

Wilcoxon Signed Rank Sum Test

rank

What is rank?

- For example, raw data were 3.2, 2.4, 5, 3.8, the ranks were 2, 1, 4, 3
- In case of ties, mid-ranks are used, e.g., if the raw data were 105, 120, 120, 121, the ranks would be 1, 2.5, 2.5, 4

W

1.5 Wilcoxon Signed Rank Sum Test

20

Subject	Drug 1	Drug 2	Diff (2-1)	Sign	Rank
1	1.9	0.7	−1.2	-	3
2	−1.6	0.8	2.4	+	8
3	−0.2	1.1	1.3	+	4.5
4	−1.2	0.1	1.3	+	4.5
5	−0.1	−0.1	0.0	NA	NA
6	3.4	4.4	1.0	+	2
7	3.7	5.5	1.8	+	7
8	0.8	1.6	0.8	+	1
9	0.0	4.6	4.6	+	9
10	2.0	3.4	1.4	+	6

Table: Hours of extra sleep on drugs 1 and 2, differences, signs and ranks of sleep study data

In the drug analysis

- Obtain S_i, the sign of $Z_i = Y_i - X_i$
- Discarding those in which $Z_i = 0$ ($S_i = 0$)
- Observations with zero differences are ignored
- Rank (R_i) = rank of $|Z_i|$ (absolute value of Z_i) after discarding $Z_i = 0$
- Signed rank: $SR = S * Rank$
- Calculate the test statistic W^+ (or W)

$$\boldsymbol{W^+ = \sum_{i=1}^{n_0} R_i I_{S_i>0}} \text{ (or } \boldsymbol{W = \sum_{i=1}^{n_0} S_i R_i}\text{)}$$

reject region

1.5 Wilcoxon Signed Rank Sum Test

- Under H_0 (no difference), W^+ could not be too small or too large
- Rejection H_0 if W^+ is too small or too large

```
> x <- c(1.9, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, 0.8, 0.0, 2.0)
> y <- c(0.7, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4)
> wilcox.test(y - x, correct = FALSE, exact = FALSE)

        Wilcoxon signed rank test

data:  y - x
V = 42, p-value = 0.02077
alternative hypothesis: true location is not equal to 0

> wilcox.test(y, x, correct = FALSE, exact = FALSE, paired = TRUE)
```

星号部分

1.5 Wilcoxon Signed Rank Sum Test*

Distribution of W^+ under H_0: $\mathrm{P}\left(W^+ = i\right) = \dfrac{t_n(i)}{2^n}, i = 0, 1, \cdots, n\left(n+1\right)/2$

$t_n(i)$: # of selecting ways to select a few numbers from 1,2, ... , n such that the sum of selected numbers equals to i .

Symmetric property: $\mathrm{P}\left(W^+ \le d\right) = \mathrm{P}\left(W^+ \ge n(n+1)/2 - d\right)$

Mean and variance: $E(W^+) = n_0(n_0+1)/4 \quad Var(W^+) = n_0(n_0+1)(2n_0+1)/24$

Rejection region $D = \{W^+ \ge c \ \ or \ \ W^+ \le d\}$ with c, d determined by type I error control

Constant c can be determined by tables and $d = \frac{n_0(n_0+1)}{2} - c$

(Exact) P-value: let w^+ be the observed value of W^+, and let $a = \max(w^+, \frac{n_0(n_0+1)}{2} - w^+)$

$$p = P[W^+ \ge a \ \text{ or } \ W^+ \le \frac{n_0(n_0+1)}{2} - a \mid H_0]$$