Contents

1 One-Sample Methods

The nonparametric tests described in this course are often called **distribution-free** test procedures because the validity of the tests does not depend on the underlying model (distribution) assumption. In other words, the significance level of the tests does not depend on the distributional assumptions.

- One-sample data: consists of observations from a single population.
- **Primary interest**: make inference about the location/center of the single distribution.
- Two popular measurements of location/center
  - Mean: sensitive to outliers
  - Median: robust against outliers
  - Mean and Median are the same for symmetric distributions.

1.1 Parametric Methods

1.1.1 One-sample Z-test (see Chapter 0.3.1)

- **Assumption**: The random sample $X_1, \cdots, X_n$ are $i.i.d.$ from $N(\mu, \sigma^2)$, where μ is the unknown mean, and σ^2 is the variance and is **known**.
- Null hypothesis $H_0 : \mu = \mu_0$ (the distribution is centered at μ_0, a prespecified null value).
- Significance level α, i.e. we want to control

$$\text{Type I error} = P(\text{Reject } H_0 | H_0 \text{ is True}) \leq \alpha.$$

- **Test statistic**

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

where $\bar{X} = 1/n \sum_{i=1}^{n} X_i$. Under H_0, $Z \sim N(0, 1)$.

Upper-tailed Z-test:

$H_a : \mu > \mu_0$

$Rejection\ Region = \{z_{obs} : z_{obs} > z_\alpha\}$

p-value $= P(Z > z_{obs}) = 1 - \Phi(z_{obs})$

Reject H_0 if z_{obs} falls into the RR or $p\text{-}value < \alpha$.

Lower-tailed Z-test:

$H_a : \mu < \mu_0$

$RR = \{z_{obs} : z_{obs} < -z_\alpha\}$

p-value $= P(Z < z_{obs}) = \Phi(z_{obs})$

Two-tailed Z-test:

$H_a : \mu \neq \mu_0$

$RR = \left\{z_{obs} : |z_{obs}| > z_{\alpha/2}\right\}$

p-value $= P(|Z| > |z_{obs}|) = 2P(Z > |z_{obs}|) = 2\{1 - \Phi(|z_{obs}|)\}$

z_α: the αth percentage point of $N(0, 1)$.

1.1.2 One-sample t-test

Suppose σ^2 is unknown, but can be estimated by the sample variance $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

- Test statistic:
$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$
- Under $H_0 : \mu = \mu_0$, $T \sim t_{n-1}$.
- **Note that the test statistic does not follow $N(0,1)$ as in the Z-test**, so the rejection region and p-value have to be calculated by using the t distribution table with $n-1$ degrees of freedom instead of standard normal table.

1.1.3 Large sample z-test

Suppose $X_1, \cdots, X_n$ are $i.i.d.$ with mean μ and variance σ^2 (both are unknown). Note here we do not assume normal distribution. Suppose n is large.

Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

By CLT, under $H_0 : \mu = \mu_0$, $Z \sim N(0,1)$ approximately for large n. So the hypothesis test can be carried out as in the z-test for normally distributed data.

Example: IQ Test

Example 1.1.1 *Ten sampled students of 18-21 years of age received special training. They are given an IQ test that is $N(100, 10^2)$ in the general population. Let μ be the mean IQ of these students who received special training. The observed IQ scores:*

$$121, 98, 95, 94, 102, 106, 112, 120, 108, 109$$

Test if the special training improves the IQ score using significance level $\alpha = 0.05$.

1. *What is the rejection region? Calculate the p-value and state your conclusion.*
2. *What if the variance is unknown?*

```
##
#### R code: analysis of the IQ score data set
##
#Test if the mean score is significantly greater than 100
#(1) suppose sigma=10 is known (z-test)
x = c(121, 98, 95, 94, 102, 106, 112, 120, 108, 109);
xbar = mean(x);
sigma=10;
n = length(x);
zval = (xbar - 100)/(sigma/sqrt(n));
zval;
pvalue = 1-pnorm(zval)

#(2) suppose sigma is unknown (one-sample t-test)
s = sd(x)
tval = (xbar - 100)/(s/sqrt(n))
tval
df=n-1
pvalue = 1-pt(tval, df)

#or use the existing R function, the default is 2-sided alternative
# test if E(X)-100>0 (alternative)
t.test(x-100, alternative ="greater")
```

1.1.4 Type I Error & Power (Z-test)

Example 1.1.2 *Refer to Example 1.1.1. Suppose $X_1, \cdots, X_n$ are $i.i.d.$ from $N(\mu, \sigma^2)$, where σ is known. Test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$. Suppose the rejection region is: $\bar{X} > 105.2$. Calculate the Type I error of this test procedure.*

In general, for an upper-tailed z-test $H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$.
For $RR = \{\bar{X} : \bar{X} > C\}$,

$$\text{Type I error} = P\left(Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{C - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{C - \mu_0}{\sigma/\sqrt{n}}\right).$$

Note: The above calculation holds approximately for the approximate z-test based on CLT when n is large.

(Upper-tailed test) $H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$. For a **level α** z-test $RR = \{\bar{X} : \bar{X} > \mu_0 + z_\alpha \sigma/\sqrt{n}\}$,

$$\text{Type I error} = P\left(Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\right) = 1 - \Phi\left(z_\alpha\right) = \alpha.$$

Power = 1-P(Type II error): the probability of detecting the departure from the null hypothesis, i.e. the chance of rejecting H_0 when H_a is true. This depends on how far the true parameter is away from the null hypothesis. We often fix the parameter value under H_a.

Example 1.1.3 *Refer to Example 1.1.1. Suppose the true mean is $\mu = 105$ ($105 > \mu_0 = 100$ so H_0 is false). Derive the power of the z-test procedure (reject H_0 when $\bar{X} > 105.2$).*

(Upper-tailed Z-test) $H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$,
$RR = \{\bar{X} : \bar{X} > C\}$. Suppose the true mean is $\mu' > \mu_0$,

$$\begin{aligned} Power(\mu') &= P(\text{Reject } H_0 | H_0 \text{ is false}) \\ &= P(\bar{X} > C | \mu = \mu') \\ &= P\left\{ Z = \frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} > \frac{C - \mu'}{\sigma/\sqrt{n}} \right\} = 1 - \Phi\left(\frac{C - \mu'}{\sigma/\sqrt{n}} \right). \end{aligned}$$

1.1.5 Confidence Interval

Suppose $X_1, \cdots, X_n$ are $i.i.d.$ from $N(\mu, \sigma^2)$ with σ known. Then $(1-\alpha)$ confidence interval for μ is:

$$[\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}].$$

Example 1.1.4 *For the IQ test example, verify that the 95% confidence interval for μ is:* $[100.3, 112.7]$.

Note that it is not correct to say that $P(100.3 < \mu < 112.7) = 0.95$. Why? In general,

$$P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}]) = 1 - \alpha,$$

where the lower and upper bounds are random variables. In the interval [100.3, 112.7], the values have been plugged in.

1.2 Binomial Test

1.2.1 Hypotheses and Test Statistic

- **Assumption**: suppose the random sample $X_1, \cdots, X_n$ are $i.i.d.$ from **a distribution with median θ**. Different from Section 1.1, here we do not require the distribution to be Normal.
- **Null hypothesis** $H_0 : \theta = \theta_0$, where θ_0 is prespecified, e.g. $\theta_0 = 0$ corresponds to testing if the distribution is centered at 0.
- Note that if the distribution is symmetric, the H_0 is equivalent to test if the population mean$= \theta_0$.
- **Alternative hypothesis** $H_a : \theta > \theta_0$ (upper-tailed test).

We consider the **binomial test statistic**

$$S = \text{number of } X_i\text{'s that exceed } \mu_0.$$

That is, S is the total number of observations out of n that exceed the hypothesized median μ_0. In another word,

$$S = \sum_{i=1}^{n} I(X_i > \mu_0),$$

where $I(A)$ is the indicator function that takes value 1 if the statement A holds and zero otherwise.

In S, we care about only the signs of $X_i - \mu_0$ (whether X_i is greater than μ_0 or not), but not the magnitudes of $X_i - \mu_0$. Therefore, the Binomial test is also called "sign test".

To determine the rejection region and calculate the p-value, we need know the distribution of the test statistic S when H_0 is true.

Distribution of S under H_0: when H_0 is true, we would expect half of the data are greater than θ_0, and half are smaller than θ_0. Therefore,

$$S \sim Binomial(n, p = 0.5) \text{ under } H_0,$$

irrespective of the underlined distribution of X_i's. Here p is the population proportion of $X_i > \theta_0$. Thus, testing $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ is equivalent to testing

$$H_0 : p = 0.5 \text{ versus } H_a : p > 0.5.$$

1.2.2 Rejection Region, Type I Error and Power

- For the upper-tailed test, a larger value of S provides more contradiction of H_0.
- Rejection region: reject H_0 when $S \geq c_{val}$, where c_{val} is the critical value that is determined to control Type I error at level α, that is, find ct such that

$$P(S \geq c_{val}|H_0) = \sum_{k=c_{val}}^{n} \binom{n}{k} 0.5^n = \alpha.$$

- Since Binomial is a discrete distribution, we may not be able to find an integer c_{val} to make the Type I error equal α exactly. In practice, we find an integer c_{val} such that the Type I error is as close to α (no larger than) as possible.

Large Sample Approximation:

For large n, by Central Limit Theorem, we know that approximately

$$\boxed{S \sim N(0.5n, 0.25n) \text{ under } H_0.}$$

Therefore, for large n, the approximate rejection region is:

$$S \geq 0.5n + z_\alpha \sqrt{0.25n},$$

which has Type I error approximately equal α.

Example 1.2.1 *Refer to Example 1.1.1. The observed IQ scores:*

$$121, 98, 95, 94, 102, 106, 112, 120, 108, 109$$

Test $H_0 : \theta = 100$ *versus* $H_a : \theta > 100$. *Suppose the rejection region is: reject when* $S \geq 8$.

1. Calculate the Type I error of this test procedure.

2. Calculate the Type I error using the Normal approximation (based on CLT).

Power calculation:

Suppose $X_1, \cdots, X_n$ are $i.i.d.$ with median θ and variance σ^2.

- $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$.
- Rejection region: $S \geq c_{val}$.
- Suppose the true median is $\theta' > \theta_0$, what is the power of the test procedure?
- Under $H_a : S \sim Binomial(n, p')$, where $p' = P(X > \theta_0) > 0.5$.

Note: We need know/assume the distribution of X_i in order to obtain p' and calculate the power.

Suppose p' is already calculated. Then the power of the sign test is:

$$\begin{aligned} \beta(\theta') &= P\{S \geq c_{val} | S \sim Binomial(n, p')\} \\ &= 1 - B(c_{val} - 1; n, p') \quad \text{(exact, small sample)} \\ &\approx 1 - \Phi\left\{\frac{c_{val} - np'}{\sqrt{np'(1-p')}}\right\} \text{(large-sample).} \end{aligned}$$

Example 1.2.2 *Suppose* $X_1, \cdots, X_{10}$ $i.i.d. \sim N(\theta, 10^2)$. *Test* $H_0 : \theta = 100$ *versus* $H_a : \theta > 100$. *Consider two tests:*
(A) z-test: reject H_0 *when* $\bar{X} > 100 + 1.645 \times 10/\sqrt{10} = 105.2$ *(Type I error: 0.05)*
(B) Binomial test: reject H_0 *when* $S \geq 8$ *(Type I error: 0.055)*

Suppose the truth is $\theta = 105$. Compare the power of the two tests for detecting such a departure from H_0. (Refer to Example 1.1.3 for the power of the z-test.)

Example 1.2.3 *Suppose X_i are $i.i.d.$ from the Laplace distribution with mean 105 and variance 100 with pdf*

$$f(x) = \frac{1}{10\sqrt{2}} \exp\left\{-\frac{|x-105|}{10/\sqrt{2}}\right\}, \quad -\infty < x < +\infty.$$

Laplace distribution is symmetric about the mean, but it has a fatter tail than the normal distribution. Find the power of the Binomial test. [Hint: $P(X > 100) = 0.753$; in R: library(VGAM); 1-plaplace(100, 105, 10/sqrt(2))]

Solution:

We know $P(X > 100) = 0.753$ when $X \sim Laplace(105, 10/\sqrt{2})$. Then under $H_a : \theta = 105$, $S \sim Binomial(n = 10, p = 0.753)$. The power of the Binomial test:

$$\beta(105) = P\{S \geq 8 | S \sim Binomial(10, 0.753)\} = 0.535.$$

Binomial test is more powerful than z-test for heavy-tailed distributions!!

Q: Can we use Table A1 to calculate the power for the above example?

Ans: Yes. Table A1 gives $P(X=k)$ for $X \sim Binomial(n,p)$ with $p \le 0.5$. If $p > 0.5$,
$P(X=k) = P\{X = n-k | X \sim Binomial(n, 1-p)\}$. So if we use table A1

$$\begin{aligned} & P\{S \ge 8 | S \sim Binomial(10, 0.753)\} \\ = \quad & P(X=0) + P(X=1) + P(X=2) \\ = \quad & 0.0586 + 0.1922 + 0.2838 = 0.5346, \end{aligned}$$

where $X \sim Binomial(10, 0.247)$.

Sample size determination: For the large sample level α sign test, $c_{val} = 0.5n + z_\alpha\sqrt{0.25n}$. In order to achieve power β, say, 90%, what is the smallest n required?

We need n such that

$$\beta(\theta') \quad \approx \quad 1 - \Phi\left\{\frac{c_{val} - np'}{\sqrt{np'(1-p')}}\right\} \geq \beta.$$

That is, we need

$$\frac{0.5n + z_\alpha\sqrt{0.25n} - np'}{\sqrt{np'(1-p')}} \leq z_\beta$$

$$\Rightarrow \quad n \geq \frac{\{1/2z_\alpha - z_\beta\sqrt{p'(1-p')}\}^2}{(0.5-p')^2} \text{ (round up).}$$

Example 1.2.4 *(Refer to Example 1.2.2) Desired power $\beta = 0.475$, calculate the minimum sample size required for the Binomial test.*

```
### R code
### Calcuate the power of the two tests using simulation.
### when Xi follow N(105,10^2)
n=10
reject.ztest = reject.binom = 0
for(j in 1:1000){
x = rnorm(n, 105, 10)
xbar = mean(x)
S = sum(x>100)
reject.ztest = reject.ztest + 1*(xbar > 105.2)
reject.binom = reject.binom + 1*(S>=8)}
reject.ztest/1000
reject.binom/1000

### when Xi follow Laplace(105,10/sqrt(2)) (variance is 100)
n=10
reject.ztest = reject.binom = 0
library(VGAM)
```

```
for(j in 1:1000){
x = rlaplace(n, location=105, scale=10/sqrt(2))
xbar = mean(x)
S = sum(x>100)
reject.ztest = reject.ztest + 1*(xbar > 105.2)
reject.binom = reject.binom + 1*(S>=8)}
reject.ztest/1000
reject.binom/1000
```

1.2.3 Hypothesis Testing: p-value

- The **p-value** is the probability of obtaining a test statistic value as extreme as the observed value, calculated assuming H_0 is true.
- Suppose $Y \sim Binomial(n, p)$, then the probability mass function of Y is:

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, \cdots, n.$$

Therefore, for the upper-tailed test, we can calculate the p-value by

$$\begin{aligned} p\text{-value} &= P(S \geq s_{obs}|H_0) \\ &= \sum_{k=s_{obs}}^{n} \binom{n}{k} (1/2)^k (1/2)^{n-k} = \sum_{k=s_{obs}}^{n} \binom{n}{k} (1/2)^n. \end{aligned}$$

The probability mass function (pmf) of Binomial distribution is tabulated in Table A1 of Higgins.

Large Sample Approximation:

For large n, $S \sim N(0.5n, 0.25n)$ approximately under H_0. So the p-value can be approximated by

$$1 - P(S \leq s_{obs} - 1|H_0) = 1 - \Phi\left\{\frac{(s_{obs} - 1) - 0.5n + 1/2}{\sqrt{0.25n}}\right\},$$

where the $1/2$ is added for continuity correction.

1.2.4 Confidence Interval for the Median θ

For a random sample $X_1, \cdots, X_n$. Sorting the random variables from the smallest to the largest results in the **order statistics** denoted as $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$. That is, $X_{(1)} = \min(X_1, \cdots, X_n)$ and $X_{(n)} = \max(X_1, \cdots, X_n)$.

The estimated median: $\hat{\theta} = X_{(\frac{n+1}{2})}$ for odd n;
$= 1/2\{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\}$ for even n. Is simple to compute and protects against outliers.

$1-\alpha$ **confidence interval:** find integers l and u such that

$$P(X_{(l)} < \theta < X_{(u)}) = 1 - \alpha.$$

- θ is the median $\Rightarrow P(X < \theta) = 0.5$.
- $\theta > X_{(l)} \Rightarrow$ at least l observations are $< \theta$.
- $\theta < X_{(u)} \Rightarrow$ at most $u-1$ observations are $< \theta$.

Let S = number of observations among n that are **less** than θ, and $S \sim Binomial(n, p = 0.5)$. That is, find l and u such that

$$\mathbf{P(l \leq S \leq u - 1) = 1 - \alpha \Leftrightarrow \sum_{k=l}^{u-1} \binom{n}{k} 0.5^n = 1 - \alpha.}$$

Note: since Binomial distribution is discrete, we may not be able to find a and b such that the equality holds exactly. In practice, we choose values l and u such that the coverage probability is close to but greater than $1 - \alpha$.

Note: unlike normal theory CI, here we use two of the order statistics of X_i to form the lower and upper bounds.

You can call the R function "conf.med" by

```
source("CImed.r")
```

For large n, $S \sim N(0.5n, 0.25n)$ approximately under H_0. Therefore, solving

$$P(l \leq S \leq u - 1) = 1 - \alpha$$

is equivalent to solving

$$\frac{l - 0.5n}{\sqrt{0.25n}} = -z_{\alpha/2}, \quad \frac{u - 1 - 0.5n}{\sqrt{0.25n}} = z_{\alpha/2}$$

for l and u and round to the nearest integers.

Example: IQ test (upper-tailed test)

Refer to Example 1.1.1. The observed IQ scores:

$$121, 98, 95, 94, 102, 106, 112, 120, 108, 109$$

Test $H_0 : \theta = 100$ versus $H_a : \theta > 100$. Significance level $\alpha = 0.05$.
(1). Calculate the p-value and conclude.

- Observed test statistic value: $s_{obs} = 7$
- Under H_0 : $S \sim Bin(10, 0.5)$
- p-value$=P(S \geq 7) = P(S = 7) + P(S = 8) + P(S = 9) + P(S = 10) = 0.1172 + 0.0439 + 0.0098 + 0.001 = 0.1719.$
- Conclusion: do not reject H_0

(3). Construct a 95% confidence interval for the median θ.

```
source("CImed.r")
x = c(121, 98, 95, 94, 102, 106, 112, 120, 108, 109)
```

```
sort(x)
# [1]  94  95  98 102 106 108 109 112 120 121
conf.med(x)   #CI for median
#median  lower  upper
 #  107     95    120
u=9; l=2
pbinom(u-1,10,0.5)-pbinom(l-1,10,0.5)
#[1] 0.9785156
```

Using the large sample approximation,
$l = 0.5*10 - 1.96\sqrt{0.25*10} = 1.9 \approx 2$ and
$u = 0.5*10 + 1.96\sqrt{0.25*10} = 8.099 \approx 8$. So the approximate 95% CI is $[X_{(2)}, X_{(8)}] = [95, 112]$.

```
pbinom(8-1,10,0.5)-pbinom(2-1,10,0.5)
#[1] 0.9345703
```

Example: IQ test (lower-tailed test)

Refer to Example 1.1.1.

- $H_0 : \theta = 120.5$ versus $H_a : \theta < 120.5$
- $\alpha = 0.05$
- Observed test statistic value: $s_{obs} = 1$.
- Under H_0 : $S \sim Bin(10, 0.5)$
- p-value$=P(S \leq s_{obs}) = P(S \leq 1) = 0.0107$
- Conclusion: reject H_0

What if we want to test $H_0 : \theta = 120.5$ versus $H_a : \theta \neq 120.5$?
Double the p-value from one-tailed test.

R code for sign test

There is no R function for sign test, so we have to do it "by hand". We use the IQ score example for illustration.

```
theta0 = 100 # hypothesised value of median
x = c(121, 98, 95, 94, 102, 106, 112, 120, 108, 109)
n = length(x)
S = sum(x > theta0)

#pbinom(y, n, p) = P(Y <= y), where Y ~ Binomial(n,p)

# for an upper tailed test
pvalue =1- pbinom(S-1, n, 1/2)
pvalue
# for a lower tailed test
pvalue = pbinom(S, n, 1/2)
pvalue
```

Advantages of Sign/Binomial Test

- Easy to implement
- Requires only mild assumptions ($X_1, \cdots, X_n$ are $i.i.d.$)
- Protects against outliers
- Efficient for heavy tailed distributions or data with outliers
- Can be used even if only signs are available

1.2.5 Inferences for Other Percentiles

- Let θ_p define the pth quantile of X, $0 < p < 1$.
- Test $H_0 : \theta_p = \theta_0$ versus $H_0 : \theta_p > \theta_0$.
- Test statistic: $S = \sum_{i=1}^{n} I(X_i > \theta_0)$.
- Under H_0: $S \sim Binomial(n, 1-p)$.
- p-value $=$
  $P(S \geq s_{obs}|H_0) = 1 - P\{S \leq s_{obs} - 1|S \sim Binomial(n, 1-p)\}$.
- Confidence interval for θ_p:

$$X_{(l)} < \theta_p < X_{(u)},$$

  - $X_{(l)}$: $(l-1)$ r.v.s will be smaller than $X_{(l)}$
  - $X_{(l)} < \theta_p \Rightarrow$ at least l of X_i's are smaller than θ_p
  - $X_{(u)} > \theta_p \Rightarrow$ at most $u-1$ of X_i's are smaller than θ_p
  - No of X_i's smaller than $\theta_p \sim Binomial(n, p)$

So l and u should satisfy

$$\sum_{k=l}^{u-1} \binom{n}{k} p^k (1-p)^{n-k} = 1 - \alpha.$$

For large n, solve

$$\frac{l - np}{\sqrt{np(1-p)}} = -z_{\alpha/2}, \quad \frac{u - 1 - np}{\sqrt{np(1-p)}} = z_{\alpha/2}.$$

Example 1.2.5 *Ex1.02 in Higgins. Exam scores:*
79 74 88 80 80 66 65 86 84 80 78 72 71 74 86 96 77 81 76 80 76 75
78 87 87 74 85 84 76 77 76 74 85 74 76 77 76 74 81 76
1. Construct a 95% CI for the 75th percentile.

Solution: The sorted data:
65 66 71 72 74 74 74 74 74 74 75 76 76 76 76 76 76 76 77 77
77 78 78 79 80 80 80 80 81 81 84 84 85 85 86 86 87 87 88 96
$n = 40$, $p = 0.75$. For $l = 25$ and $u = 36$:

```
> pbinom(u-1, n, p)-pbinom(l-1,n,p)
[1] 0.9577129
```

So CI: $(X_{(25)}, X_{(36)}) = (80, 86)$. Or using the normal approximation: $np = 30$, $\sqrt{np(1-p)} = 2.74$,

$$\frac{l-30}{2.74} = -1.96 \Rightarrow l = 24.6 \approx= 25$$

$$\frac{u-1-30}{2.74} = 1.96 \Rightarrow u = 36.37 \approx= 36$$

CI: $(X_{(25)}, X_{(36)}) = (80, 86)$.

2. Test whether the 75th percentile is greater than 79.

$S = 16, P\{S >= 16 | S \sim B(40, 0.25)\} = 1 - pbinom(15, 40, 0.25) =$ 0.026. So the 75th percentile is significantly greater than 79.

1.3 Wilcoxon Signed Rank Test

1.3.1 Hypothesis Testing

The Binomial (sign) test utilizes only the signs of the differences between the observed values and the hypothesized median. We can use the signs as well as the ranks of the differences, which leads to an alternative procedure.

Basic Assumptions:

- $X_1, \cdots, X_n$ are independently and identically distributed, continuous and symmetric about the common median θ.
- Test $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$, where θ_0 is a prespecified hypothesized value.

Steps for calculating the Wilcoxon signed rank statistic:

- Define $D_i = X_i - \theta_0$ and rank $|D_i|$.

- Define R_i as the rank of $|D_i|$ corresponding to the ith observation.
- Let
$$S_i = \begin{cases} 1 & \text{if } D_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$
- Define the Wilcoxon signed rank statistic as
$$SR_+ = \sum_{i=1}^{n} S_i R_i,$$

the sum of ranks of positive D_i's, i.e. the positive signed ranks.

- If H_0 is true, the probability of observing a positive difference D_i of a given magnitude is equal to the probability of observing a negative difference of the same magnitude.
- Under H_0, the sum of positive signed ranks is expected to have the same value as the sum of negative signed ranks.

- Thus a large or a small value of SR_+ indicates a departure from H_0.
- For the two-sided test $H_a : \theta \neq \theta_0$, we reject H_0 if SR_+ is either too large or too small.
- For upper tailed test $H_a : \theta > \theta_0$, we reject H_0 if SR_+ is too large.
- For lower tailed test $H_a : \theta < \theta_0$, we reject H_0 if SR_+ is too small.

Ties The assumed continuity in the models implies there are no ties. However, often there are ties in practice. When there are ties, the ranks will be replaced by the midranks of the tied observations. If there are zero D_i values, discard them and then recalculate the ranks.

Example 1.3.1 *Cooper et al. (1967). Objective: to test the hypothesis of no change in hypnotic susceptibility versus the alternative that hypnotic susceptibility can be increased with training. Hypnotic susceptibility is a measurement of how easily a person can be hypnotized. (Pair replicate data)*

Table 1: Average Scores of Hypnotic Susceptibility

Subject	X_i (before)	Y_i (after)	$D_i = Y_i - X_i$	R_i	S_i
1	10.5	18.5	8	6	1
2	19.5	24.5	5	5	1
3	7.5	11	3.5	4	1
4	4.0	2.5	-1.5	2.5	0
5	4.5	5.5	1	1	1
6	2.0	3.5	1.5	2.5	1

Let θ be the median of D_i, the differences. H_0: $\theta = 0$ versus $H_a : \theta > 0$.

- Calculate $D_i = 8, 5, 3.5, -1.5, 1, 1.5$.
- Obtain R_i and S_i
- $SR_+ = 18.5$ seems big, suggesting maybe HS has increased. But is 18.5 large enough to reject H_0? How can we be more precise in our conclusion?

Critical values for the Wilcoxon Signed Rank Test Statistic and p-value.

To obtain p-value of a test, we must know the distribution of the test statistic under H_0. Under H_0,

- D_i or $-D_i$ are equally likely;
- R_i is equally likely to have come from a positive or negative D_i;
- For each subject i, R_i has $1/2$ chance to be included in SR_+. So there are total 2^n equally likely outcomes, since each of n R_i's may or may not be included in SR_+.
- In Example 1.3.1, $n = 6$ so total $2^6 = 64$ outcomes.

Table 2: The distribution of SR_+ under H_0.

Ranks Included	SR_+	Probability $P(SR_+ = t)$
None	0	1/64
1	1	1/64
2	2	1/64
..	..	..
6	6	1/64
1,2	3	1/64
1,3	4	1/64
..	..	..
1,2,3,4,5,6	21	1/64

Therefore, p-value=$\sum_{t:t\geq 18.5} P(SR_+ = t) = 0.063$.

Table A9: critical values for SR_+ assuming no zero and ties.

TABLE A9
Signed-Rank Tail Probabilities, $P(SR_+ \geq c)$

c	$n = 4$
7	.250
8	.188
9	.125
10	.063

c	$n = 5$
10	.250
11	.188
12	.125
13	.094
14	.063
15	.031

c	$n = 6$
14	.234
15	.172
16	.125
17	.094
18	.063
19	.047
20	.031
21	.016

c	$n = 7$
18	.242
19	.195
20	.156
21	.117
22	.086
23	.063
24	.047
25	.031
26	.023
27	.016
28	.008

c	$n = 8$
23	.234
24	.195
25	.160
26	.129
27	.102
28	.078
29	.059
30	.043
31	.031
32	.023
33	.016
34	.012
35	.008
36	.004

c	$n = 9$
28	.250
29	.215
30	.182
31	.152
32	.127
33	.104
34	.084
35	.066
36	.051
37	.039
38	.029
39	.021
40	.016
41	.012
42	.008
43	.006
44	.004
45	.002

c	$n = 10$
34	.247
35	.217
36	.188
37	.162
38	.139
39	.117
40	.098
41	.081
42	.066
43	.054
44	.043
45	.033
46	.025
47	.020
48	.015
49	.011
50	.008
51	.006
52	.004
53	.003
54	.002
55	.001

c	$n = 11$
41	.233
42	.207
43	.183
44	.161
45	.140
46	.121
47	.104
48	.088
49	.074
50	.062
51	.051
52	.042
53	.034
54	.027
55	.021
56	.017
57	.013
58	.010
59	.007
60	.005
61	.004
62	.003
63	.002
64	.001
65	.001
66	.000

c	$n = 12$
48	.235
49	.212
50	.190
51	.170
52	.151
53	.133
54	.117
55	.102
56	.088
57	.076
58	.065
59	.055
60	.046
61	.039
62	.032
63	.026
64	.021
65	.017
66	.014
67	.011
68	.008
69	.006
70	.005
71	.004
72	.003
73	.002
74	.001
75	.001
76	.001
77	.000
78	.000

Lower-tail probabilities may be obtained as $P(SR_+ \leq c) = P[SR_+ \geq n(n+1)/2 - c]$.

The critical values of the Wilcoxon Signed Rank test statistic are tabulated for various sample sizes; see Hollander and Wolfe (1999). The tables of exact distribution of SR_+ based on permutations is given in Higgins (2004).

In R, the function wilcox.test(z,alternative="greater")

gives the test statistic value SR_+ and the p-value for two-sided/lower/upper-tailed tests.

```
> D=c(8,5,3.5,-1.5,1,1.5)
> wilcox.test(D, alternative="greater")
Wilcoxon signed rank test with continuity correction
V = 18.5, p-value = 0.05742
alternative hypothesis: true location is greater than 0
```

The V $= 18.5$ here is SR_+.

Normal Approximation

It can be shown that for large sample n, the null distribution of SR_+ is approximately normal with mean μ and variance σ^2 where

$$\mu = \frac{n(n+1)}{4}, \quad \sigma^2 = \frac{n(n+1)(2n+1)}{24}.$$

So the Normal cut-off points can be used for large values of n.

When there are ties, use the mid-ranks for R_i. For large sample procedure we must use a different variance

$$Var_0(SR_+) = \frac{1}{24}\left\{n(n+1)(2n+1) - 1/2\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)\right\},$$

where g=# of tied groups of $|D_i|$, t_j=size of the jth tied group.

Derivation of the mean & variance of SR_+ under H_0 (optional)

Define independent variables $V_1, V_2, \cdots, V_n$, where $P(V_i = i) = P(V_i = 0) = 1/2$, $i = 1, \cdots, n$. That is, either V_i is a rank (when sign is positive) or it is zero (when sign is negative). Under H_0, the two probabilities are equal. Then under H_0, SR_+ has the same distribution as $\sum_{i=1}^n V_i$ (like summing up the ranks).

$$E(V_i) = i(1/2) + 0(1/2) = i/2, \quad E(V_i^2) = i^2(1/2) + 0^2(1/2) = i^2/2,$$

$$Var(V_i) = E(V_i^2) - \{E(V_i)\}^2 = i^2/2 - (i/2)^2 = i^2/4.$$

$$E_0(SR_+) = E(\sum_{i=1}^n V_i) = \sum_{i=1}^n E(V_i) = \frac{1}{2}\sum_{i=1}^n i = \frac{1}{2}\left\{\frac{n(n+1)}{2}\right\} = \frac{n(n+1)}{4}.$$

Due to independence of V_i,

$$\begin{aligned} Var_0(SR_+) &= Var(\sum_{i=1}^{n} V_i) = \sum_{i=1}^{n} Var(V_i) = \frac{1}{4}\sum_{i=1}^{n} i^2 \\ &= \frac{1}{4}\left\{\frac{n(n+1)(2n+1)}{6}\right\} = \frac{n(n+1)(2n+1)}{24}. \end{aligned}$$

Note that these formulas hold exactly for any n.

If there are ties, then let V_i equal either 0 or the midrank assigned to observation i, each with probability 1/2. Then $E_0(SR_+) = 1/2$(sum of midranks)$=n(n+1)/4$ and $Var_0(SR_+) = 1/4$(sum of squared midranks)$\leq n(n+1)(2n+1)/24$.

Example 1.3.2 *The following data gives the depression scale factors of 9 patients, measured before therapy (X) and after therapy* Y.

i	1	2	3	4	5	6	7	8	9
X_i	1.83	0.5	1.62	2.48	1.68	1.88	1.55	3.06	1.30
Y_i	0.88	0.65	0.59	2.05	1.06	1.29	1.06	3.14	1.29
$X_i - Y_i$	0.95	-0.15	1.03	0.43	0.62	0.59	0.49	-0.08	0.01

Carry out the Signed Rank test to test if the therapy reduces the depression. Use significance level $\alpha = 0.05$.

Example 1.3.3 *Suppose* D_i *are -3, -2, -2, -1, 1, 1, 3, 4. Test* $H_0 : \theta = 0$ *versus* $H_a : \theta \neq 0$. *Compare the null variances of* SR_+ *with and without correction for ties.*

1.3.2 Hodges-Lehmann Estimator of θ

Estimation of θ: the center of symmetry of the X_i's (one sample).

- Form the $n(n+1)/2$ "Walsh" averages

$$W_{ij} = (X_i + X_j)/2, \quad 1 \leq i \leq j \leq n.$$

  Include the X_i's themselves (i.e. when $i = j$). Do not discard any $X_i = 0$.

- The **Hodges-Lehmann's estimator**

$$\widehat{\theta} = \text{median}_{1 \leq i \leq j \leq n}(W_{ij}).$$

Example 1.3.4 *Refer to Hypnotic Susceptibility Example 1.3.1. Estimate the treatment effect θ.*

Solution: $D_i = Y_i - X_i = 8, 5, 3.5, -1.5, 1, 1.5$. The treatment effect θ is the center of D_i. $n = 6$ so total 21 Walsh averages.

Walsh Avg.	8	5	3.5	-1.5	1	1.5
8	8	6.5	5.75	3.25	4.5	4.75
5		5	4.25	1.75	3	3.25
3.5			3.5	1	2.25	2.5
-1.5				-1.5	-0.25	0
1					1	1.25
1.5						1.5

The sorted 21 Walsh averages are:
-1.50 -0.25 0.00 1.00 1.00 1.25 1.50 1.75 2.25 2.50 3.00 3.25 3.25 3.50 4.25 4.50 4.75 5.00 5.75 6.50 8.00.

Therefore, the median of the 21 Walsh averages is $\widehat{\theta} = 3$.

The calculation is tedious by hand.

```
R code:
x=c(10.5,19.5,7.5,4.0,4.5,2.0)
y=c(18.5,24.5,11,2.5,5.5,3.5)
print(d <- y - x)
walsh <- outer(d, d, "+") / 2
walsh <- walsh[!lower.tri(walsh)]
sort(walsh)
median(walsh)
```

Miscellous Properties of HL's Estimator:

- For symmetric distributions, $\widehat{\theta}$ is a consistent estimator for θ
- $\widehat{\theta}$ is insensitive to outliers

Relationship of HL's Estimator with Signed-Rank-Test:

Null hypothesis $H_0 : \theta = \theta_0$, where θ is the median of X_i. Let

$$W^+ = \#(W_{ij} > \theta_0).$$

When there are no ties or zero D_i's, $W^+ = SR_+$. Therefore, we can also perform test using Walsh's averages.

(For the Hypnotic example, $W^+ = 18$, $SR_+ = 18.5$. The slight difference is due to ties).

HL's confidence Interval for θ.

Construction of a symmetric two-sided $(1-\alpha)$ confidence interval:

- Obtain the upper $(\alpha/2)$th percentage point of the null distribution of SR_+ from Table, denoted by $sr_{+,\alpha/2}$
- Set $C_\alpha = \frac{n(n+1)}{2} + 1 - sr_{+,\alpha/2}$
- Let $W^{(1)} \leq \cdots \leq W^{(M)}$ denote the ordered Walsh's averages, $M = n(n+1)/2$.

- The $(1-\alpha)$ confidence interval for θ is given by

$$[\theta_L = W^{(C_\alpha)}, \theta_U = W^{(M+1-C_\alpha)}].$$

- This interval contains all the "acceptable" θ_0's so that $H_0: \theta = \theta_0$ is not rejected by the small sample 2-sided signed rank test with significance level α.

For large sample size n, the integer C_α can be approximated by

$$C_\alpha \approx \frac{n(n+1)}{4} - z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}} \text{ (round down).}$$

Example 1.3.5 *Construct a 90% confidence interval for the treatment effect in the Hypnotic example 1.3.1.*

Example 1.3.6 *Construct a 90% confidence interval for median hypnotic susceptibility before training in the Hypnotic example 1.3.1. Data: 10.5,19.5,7.5,4.0,4.5,2.0*

Walsh Avg.	10.5	19.5	7.5	4.0	4.5	2.0
10.5	10.5	15	9	7.25	7.5	6.25
19.5		19.5	13.5	11.75	12	10.75
7.5			7.5	5.75	6	4.75
4				4	4.25	3
4.5					4.5	3.25
2.0						2

W_{ij} : 10.50 15.00 19.50 9.00 13.50 7.50 7.25 11.75 5.75 4.00 7.50 12.00 6.00 4.25 4.50 6.25 10.75 4.75 3.00 3.25 2.00