Alternatives to the t-tests

Professor: Hammou El Barmi
Columbia University

The goal of this chapter is

- Distinguish parametric and nonparametric test procedures
- Explain commonly used nonparametric test procedures
- Perform hypothesis testing using nonparametric procedures

Parametric test procedures:

- involve population parameters
- have stringent assumptions 有严格的假设
- perform hypothesis tests using nonparametric procedures 这句话怎么理解？

Nonparametric test procedure

- make less stringent demands of the data.
- have stringent assumptions
- data can be measured on any scale
- results may be as exact as parametric procedures

Inference using t-distribution: one sample case

- This chapter introduces some nonparametric techniques we use when you don't believe the assumptions for the stuff you learned before.
- All procedures that you learned before come in natural triples
  - a hypothesis test
  - a confidence interval obtained by inverting the test,
  - the point estimate obtained by shrinking the confidence level to zero. This is called the Hodges-Lehmann estimator associated with the confidence interval.
- A familiar example is
  - hypothesis test: t test
  - confidence interval: t confidence interval
  - point estimate: sample mean

Now we are going to learn about some competing techniques

- hypothesis test: sign test
- confidence interval: associated confidence interval
- point estimate: sample median

and

- hypothesis test: Wilcoxon signed rank test
- confidence interval: associated confidence interval
- point estimate: associated Hodges-Lehmann estimator

- $Y_1, Y_2, \ldots, Y_n$ a random sample from $N(\mu, \sigma^2)$
- $\bar{Y}$ sample mean and s is the sample standard deviation.
- Goal: test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$
- Test statistics

  $$t = \frac{\bar{Y} - \mu_0}{SE(\bar{Y})}$$

  根据中心极限定理，样本均值服从总体均值为期望，总体单位差除以样本量为单位差的一个正态分布。

  where

  $$SE(\bar{Y}) = s/\sqrt{n}$$

  is the standard error of $\bar{Y}$.
- Reject H_0 if $|t| > t_{n-1}(\alpha/2)$ or if p-value $< \alpha$
- If $H_a : \mu > \mu_0$, reject H_0 if $t > t_{n-1}(\alpha)$ or if p-value $< \alpha$
- If $H_a : \mu < \mu_0$, reject H_0 if $t < -t_{n-1}(\alpha)$ or if p-value $< \alpha$
- A $100(1-\alpha)\%$ confidence interval for μ is

  $$\bar{Y} \pm t_{n-1}(\alpha/2) SE(\bar{Y})$$

Inference using t-distribution: two sample case

- $Y_{11}, Y_{12}, \dots, Y_{1n_1}$ a random sample from $N(\mu_1, \sigma_1^2)$
- $Y_{21}, Y_{22}, \dots, Y_{2n_2}$ a random sample from $N(\mu_2, \sigma_2^2)$
- $\bar{Y}_1$ and s_1^2 sample mean and variance of the first sample
- $\bar{Y}_2$ and s_2^2 sample mean and variance of the first sample
- Goal: test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$
- Two cases to consider
    - $\sigma_1^2 = \sigma_2^2$ (Equal Variances Case)
    - $\sigma_1^2 \neq \sigma_2^2$ (Unequal Variances Case)

- Case 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{SE(\bar{Y}_1 - \bar{Y}_2)}$$

where

$$SE(\bar{Y}_1 - \bar{Y}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

with

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}},$$

the pooled estimate of the standard deviation σ.

- Reject H_0 if $|t| > t_{n_1+n_2-2}(\alpha/2)$ or if p-value $< \alpha$
- If $H_a : \mu > \mu_0$, reject H_0 if $t > t_{n_1+n_2-2}(\alpha)$ or if p-value $< \alpha$
- If $H_a : \mu < \mu_0$, reject H_0 if $t < -t_{n_1+n_2-2}(\alpha)$ or if p-value $< \alpha$
- A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{n_1+n_2-2}(\alpha/2)SE(\bar{Y}_1 - \bar{Y}_2)$$

- Case 1: $\sigma_1^2 \neq \sigma_2^2$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- t does not follow a t-distribution under H_0. However, its distribution can be well approximated by a t-distribution with degrees of freedom equal to the following complicated formula:

$$d = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

- approximation is known as Satterthwaite's Method
- Reject H_0 if $|t| > t_d(\alpha/2)$ or if p-value $< \alpha$
- If $H_a : \mu > \mu_0$, reject H_0 if $t > t_d(\alpha)$ or if p-value $< \alpha$
- If $H_a : \mu < \mu_0$, reject H_0 if $t < -t_d(\alpha)$ or if p-value $< \alpha$
- A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{Y}_1 - \bar{Y}_2 \pm t_d(\alpha/2) SE(\bar{Y}_1 - \bar{Y}_2)$$

- Let $X_1, X_2, \ldots, X_n$ be a random sample from some continuous distribution F
- The continuity assumption assures that ties are impossible. With probability one we have. 连续性假设怎么理解？为什么在假设检验中对随机变量的分布需要有连续性假设？
- The continuity assumption is necessary for exact hypothesis tests.
- The continuity assumption is unnecessary for the point estimate and confidence interval.
- The parameter of interest is $\eta = \text{median(F)}$
- Because of the continuity assumption, the median is uniquely defined.

Sign Test for a population median

- Goal test $H_0 : \eta = \eta_0$ against $H_a : \eta \neq \eta_0$
- Let

$$
\begin{aligned}
T &= \sum_{i=1}^{n} I(X_i > \eta_0) \\
&= \sum_{i=1}^{n} s(X_i - \eta_0)
\end{aligned}
$$

where $s(x) = 1$ if $x > 0$ and $s(x) = 0$ otherwise. T counts the numbers of X_is that exceed η_0. 因为如果假设成立，则对于每一个观测值，他都以1/2的概率可能被计算为1。每一个数据都是一

- If H_0 is true then $T \sim Bin(n, 1/2)$ (this implies that $E(T) = n/2$ and $Var(T) = n/4$) 伯努利分布。这里一共有N个数据，所以T表示的是着N个数据的和。是一个二项分布
- Reject H_0 is T is much less than or much greater than $n/2$. That is if $|T - n/2|$ is large
- The p-value $= 2\min(P(T \leq t), P(T \geq t))$ where t is the realized value of T based on the sample
- To generate a $100(1-\alpha)\%$ confidence interval for the true median, use the values of η for which we fail to reject H_0 at α.

将原假设中的t带入数据可以得到一个具体的t值。又因为T这个随机变量是服从二项分布的，所以可以计算出他的具体概率。选择小的那个是因为他描述了一种与他相比距离期望值更远的概率区间。这样的区间是符合P值的定义的。

- In R, install the package BSDA
- To test that median is 1 against the median not zero at 5% use

```
SIGN.test(data, md=1,alternative="two.sided",conf.level=0.95)
```

- One sided test possible

```
SIGN.test(data, md=1,alternative="greater",conf.level=0.95)
SIGN.test(data, md=1,alternative="less",conf.level=0.95)
```

Example (Income data):

```
income <- c(7, 1110, 7.1, 5.2, 8, 12, 0, 5, 2.1, 2, 46, 7.5)
```

Figure: Histogram and Box Plot

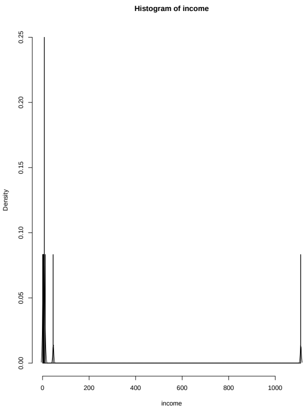


Figure: Box plot

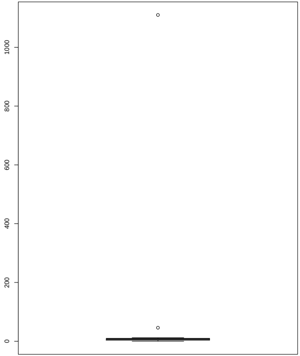
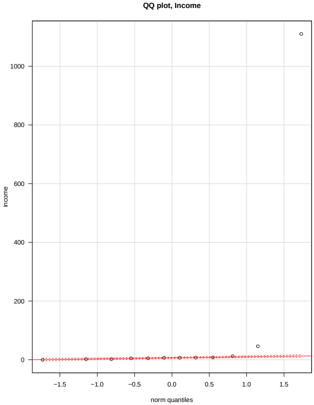


Figure: QQ plot



```
> summary(income)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   0.000    4.275    7.050  101.000    9.000 1110.000

> t.test(income, mu=1)

One Sample t-test

data:  income
t = 1.0893, df = 11, p-value = 0.2993
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 -101.0454  303.0287
sample estimates:
mean of x
 100.9917
```

- Presence of outliers has a dramatic effect on a 95% confidence interval for the population mean μ which is $[-101, 303]$
- This t based confidence interval is suspect because the normality assumption is unreasonable.
- A confidence interval for the population median income is more sensible because the median is more likely to be a more reasonable typical value.
- Using the sign procedure, a 95% confidence interval for the population median is $[2.32, 11.57]$.

当数据具有很大的偏度的时候，均值对数据中心的代表性弱于中位数。因为T检验就不是很合理了。因为

```
> SIGN.test(income, md=1, alternative="two.sided")

One-sample Sign-Test

data:  income
s = 11, p-value = 0.006348
alternative hypothesis: true median is not equal to 1
95 percent confidence interval:
  2.408455 11.574545
sample estimates:
median of x
       7.05

                  Conf.Level L.E.pt  U.E.pt
Lower Achieved CI     0.8540 5.0000  8.0000
Interpolated CI       0.9500 2.4085 11.5745
Upper Achieved CI     0.9614 2.1000 12.0000
> 2*(1-pbinom(10,12,0.5))
[1] 0.006347656
```

```
> age<-c(54, 42, 51, 54, 49, 56, 33, 58, 54, 64, 49)
> sort(age)
 [1] 33 42 49 49 51 54 54 54 56 58 64

> summary(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  33.00   49.00   54.00   51.27   55.00   64.00
```

Question of interest: is the typical age at first transplant 50?

Figure: Histogram and Box Plot

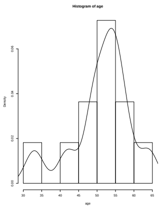


Figure: Box plot

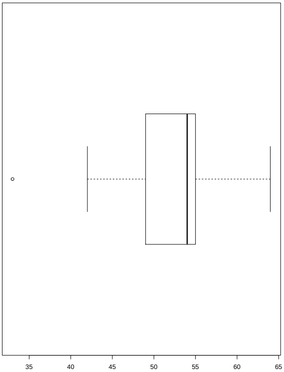
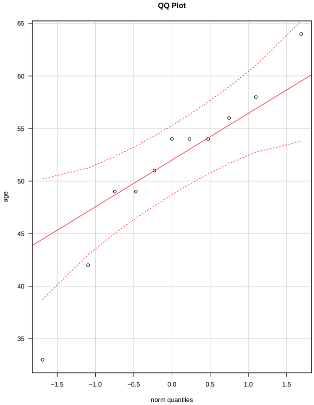


Figure: QQ plot



- The normal QQ-plot of the sample data indicates mild deviation from normality in the left tail
- It is good practice in this case to use the nonparametric test as a double-check of the t-test.

```
> t.test(age, mu=50)

One Sample t-test

data:  age
t = 0.51107, df = 10, p-value = 0.6204
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 45.72397 56.82149
sample estimates:
mean of x
 51.27273
```

A 95% confidence for the mean in [45.72, 56.82]

```
> SIGN.test(age, md=50)

One-sample Sign-Test

data:  age
s = 7, p-value = 0.5488
alternative hypothesis: true median is not equal to 50
95 percent confidence interval:
 46.98909 56.57455
sample estimates:
median of x
         54

                  Conf.Level  L.E.pt  U.E.pt
Lower Achieved CI     0.9346 49.0000 56.0000
Interpolated CI       0.9500 46.9891 56.5745
Upper Achieved CI     0.9883 42.0000 58.0000
```

A 95% confidence for the mean in [46.99, 56.59]. Similar conclusion was reached with the t confidence interval and the t-test on μ. You should have less confidence in these results because the normality assumption is tenous.

N(0,1)

- The following result holds. Under H_0

$$\lim_{n\to\infty} P\left(\frac{S-n/2}{\sqrt{n/4}} \le x\right) = \Phi(x)$$

- A large sample test reject H_0 if

$$S < n/2 - Z_{\alpha/2}\sqrt{n}/2 \quad \text{or} \quad S < n/2 + Z_{\alpha/2}\sqrt{n}/2$$

- How do you do a one sided test in this case?

检验是否对称分布

- The Wilcoxon procedure assumes you have a random sample $X_1, X_2, \ldots, X_n$ from a symmetric distribution (need not be normal) $\Rightarrow$ mean $=$ median
- To test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ requires
  - sign of $X_i - \mu_0$
  - ranks $R_1, R_2, \ldots, R_n$ of the $|X_i - \mu_0|$s
- The test statistic is

$$W = \sum_{i=1}^{n} s(i) R_i$$

  where $s(i) = 1$ if $X_i - \mu_0 > 0$ and 0 otherwise.
- Under H_0

$$E(W) = \frac{n(n+1)}{4}$$

1/2 * sum(Ri)

  and we reject H_0 if $|W - n(n+1)/4|$ is large.
- in R use wilcox.test(data)

var(W) = sum(j^2 * var(wj))
= n(n+1)(2n+1)/24

Example: here $\mu_0 = 10$.

X_i	X_i-10	sign	$\lvert X_i - 10 \rvert$	rank	sign $\times$ rank
20	10	$+$	10	6	6
18	8	$+$	8	4.5	4.5
23	13	$+$	13	8	8
5	-5	-	10	3	-3
14	4	$+$	10	4	4
8	-2	-	2	1	-1
18	8	$+$	8	4.5	4.5
22	12	$+$	12	7	7

这个Rank是有问题。就是排序

$$W = 6 + 4.5 + 8 + \overset{4}{2} + 4.5 + 7 = 32$$ 只取正数

$n = 8$ this implies that under $H_0, E(W) = 8(9)/2 = 18$

```
> wilcox.test(x,mu=10,conf.int=TRUE,correct=FALSE)

Wilcoxon signed rank test

data:  x
V = 32, p-value = 0.04967
alternative hypothesis: true location is not equal to 10
95 percent confidence interval:
 10.99996 21.00005
sample estimates:
(pseudo)median
      16.0056
```

用于成对数据的检验

- Example: Two remedies for insomnia and the number of hours of sleep gained was recorded. Same people used in the experiment

```
a<-c(0.7, -1.6, -0.2, -1.2, 0.1, 3.4, 3.7, 0.8, 0.0, 2.0)
b<- c(1.9, 0.8, 1.1, 0.1,  -0.1, 4.4, 5.5, 1.6, 4.6, 3.0)
```

- Goal: test $H_a : \mu_a = \mu_b$ against $H_a : \mu_a \neq \mu_b$

```
d<-a-b
> sleep<-data.frame(a , b, d)
> shapiro.test(sleep$d)
Shapiro-Wilk normality test

data:  d
W = 0.83798, p-value = 0.04173

p-value< 0.05, we reject the null hypothesis that the differences are
from a normal distribution.
```

```
> wilcox.test(d, mu=0, conf.int=TRUE)

Wilcoxon signed rank test with continuity correction

data:  d
V = 1, p-value = 0.008004
alternative hypothesis: true location is not equal to 0
95 percent confidence interval:
 -2.7999620 -0.7999339
sample estimates:
(pseudo)median
     -1.299966
```

- The following result holds. Under H_0

$$\lim_{n\to\infty} P\left(\frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \leq x\right) = \Phi(x)$$

- A large sample test reject H_0 if

$$S < n(n+1)/4 - Z_{\alpha/2}\sqrt{n(n+1)(2n+1)/24}$$

or if

$$S > n(n+1)/4 + Z_{\alpha/2}\sqrt{n(n+1)(2n+1)/24}$$

- How do you do a one sided test in this case?

- For symmetric distributions, the t, the sign and Wilcoxon procedures are appropriate
- It the underlying distribution is extremely skewed, you can use the sign procedure to get a confidence interval for the population median
- Otherwise, you can transform the data to a scale where the underlying distribution is nearly normal and use t .
- Moderate degrees of skewness will not have a big impact on the t based results.
- Data from heavy-tailed distribution can have a profound impact on the t-test and the t confidence interval
- The sign and the Wilcoxon procedures downweight the influence of outliers by looking at sign and signed ranks instead of the actual values.
- A weakness of nonprametric methods is that they do not easily generalize to complex problems

- The WMW procedure assumes you have independent random samples from two populations
- Assumes the populations have the same shapes and spreads. Their distributions are not required to be symmetric
- The WMW procedure gives a confidence interval for the difference $\eta_1 - \eta_2$, the difference of the populations' medians.

- $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent random samples from two populations
- This test trie to detect location shifts
- To compute the test statistic (see below), combine the two samples and rank the observations.
- Let T_X be the sum of the ranks for the observations in the X-group and T_Y for the Y-group (note that $T_X + T_y = (n_1 + n_2)(n_1 + n_2 + 1)/2.$)
- Under $H_0 : \eta_1 = \eta_2, E(T_X) = n_1(n_1 + n_2 + 1)/2$ and $Var(T_X) = n_1 n_2(n_1 + n_2 + 1)/12$
- In some applications the Mann-Whitney form U_X of Wilcoxon rank sum test is used:

$$U_X = T_X - \frac{n_1(n_1+1)}{2}$$

As such

$$U_X = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i > Y_j)$$

- The test is most appropriate when the populations have the same shape and differ only in location (same in dispersion). The distributions do not have to be symmetric.
- $$\lim_{\min(n_1,n_2)\to\infty} P\left(\frac{T_X - n_1(n_1+n_2+1)/2)}{\sqrt{n_1 n_2(n_1+n_2+1)/12}} \leq x\right) = \Phi(x)$$

(Wilcoxon) Mann-Whitney two sample procedure

```
group1 (X)<-c(0.8, 2.8, 4.0, 2.4, 1.2, 0.0, 6.2, 1.5, 28.8, 0.7)
group2 (Y) <-c(2.3, 0.3, 5.2, 3.1, 1.1, 0.9, 2.0, 0.7, 1.4, 0.3)

Wilcoxon rank sum test with continuity correction

data:  x and y
W = 61.5, p-value = 0.4053
alternative hypothesis: true location shift is not equal to 0
```