MAT 5030
Chapter 5:
One and two-sample tests

Kazuhiko Shinki

Wayne State University

Introduction

Recall that ***hypothesis testing*** is a decision making rule between a null hypothesis H_0 and an alternative hypothesis H_1.

Usually the two hypotheses are NOT symmetric. H_0 is rejected only when there is a strong evidence that H_0 is unlikely. H_0 is accepted otherwise.

Example:

We found that 51 out of 100 newborn babies are male and want to test

- "H_0 : 50% of babies are male in U.S." against
- "H_1 : Not 50% of babies are male in U.S.".

Due to the small sample size, we probably do not reject H_0 (so do accept H_0) but still neither support H_0 nor H_1.

5.1 One-sample t-test

Example: Math PhD Salary

The following figures are starting salary for 7 math PhDs:
\$45,000, \$80,000, \$50,000, \$95,000, \$65,000, \$110,000, \$48,000.

Suppose sallaries follow a normal distribution. Test if the average starting salary μ for math PhDs is significantly different from \$79,500. (For example, \$79,500 is the average salary for all PhDs in US).

- $H_0 : \mu = 79,500$ vs $H_1 : \mu \neq 79,500$.

Idea: Reject H_0 if $(\bar{x} - 79,500)$ is far from zero. The question is what is "far from".

5.1 One-sample t-test

Theorem:

Let $x_1, \cdots, x_n$ are observations from independent and identically distributed (***IID***) normal distribution $N(\mu_0, \sigma^2)$. Then,

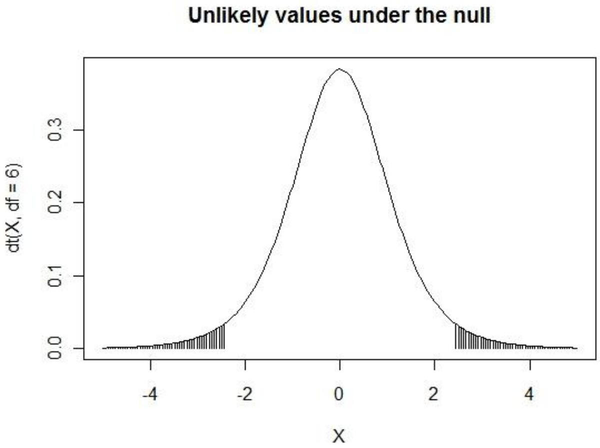
$$T := \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1) \tag{1}$$

Note:

- "$\sim$" means the right hand side follows the distribution on the left hand side. $t(k)$ is t-distribution with degrees of freedom k.
- T is called ***t-statistic*** .

5.1 One-sample t-test

Reject H_0, if the realization of $|T|$ is extremely large for $t(n-1)$. Typically reject H_0 when T falls in the most extreme 5% of the range (called ***rejection region*** with a 5% significance level).



Note: code for the previous figures

```
# 1st figure (slides no.9)
X <- 0.05*(-100:100)
plot(X, dt(X, df=6),type="l",xlim=c(-5,5))
lines(c(-0.9465,-0.9465), c(0,dt(-0.9465,df=6)))

# 2nd figure (slides no.5)
plot(X, dt(X, df=6),type="l",xlim=c(-5,5),
  main="Unlikely values under the null")
Q <- qt(0.975,df=6) # Q = 2.4469
SQ <- seq(Q, 5, length = 50)
for (i in 1:50){
lines(c(SQ[i],SQ[i]), c(0,dt(SQ[i],df=6)))
lines(c(-SQ[i],-SQ[i]), c(0,dt(-SQ[i],df=6)))
}
```

5.1 One-sample t-test

Go back to our example:

If $H_0 : \mu = 79,500$ is true, the sample mean salary follows $t(7-1) = t(6)$. The actually sample mean is

$$T = \frac{70428.57 - 79500}{25356.51/\sqrt{7}} = -0.94652.$$

Want to see if this number is an usual realization of $t(6)$.

5.1 One-sample t-test

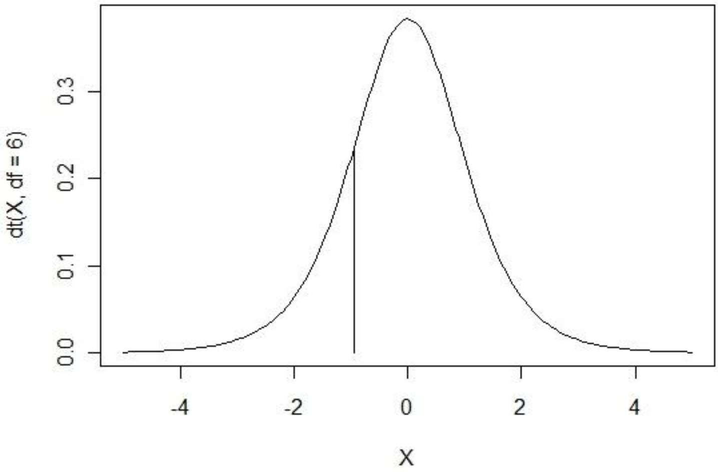
Rejection region:

If we set the upper and lower 2.5% each of the total area as the rejection region, we reject $\boldsymbol{H_0}$ if $\boldsymbol{T > 2.447}$ or $\boldsymbol{T < -2.447}$ (namely $\boldsymbol{|T| > 2.447}$) since

```
> qt(0.975,df=6)
[1] 2.446912
> qt(0.025,df=6)
[1] -2.446912
```

Since $\boldsymbol{|T| = |-0.94652| < 2.447}$, $\boldsymbol{H_0}$ is not rejected. Math PhDs' starting salary is NOT significantly different from all PhDs'.

5.1 One-sample t-test



Notation for t random variables:
Let T be a t-random variable with $DF = n$. Define $t_\alpha(n)$ ($0 < \alpha < 1$) be the constant such that

$$P(T \leq t_\alpha(n)) = \alpha.$$

Note:

- $-t_\alpha(n) = t_{1-\alpha}(n)$ because t-distribution is symmetric around 0.
- $t_\alpha(n)$ depends on the degrees of freedom n, but we often omit n.
- Notation z_α is used for a standard normal distribution.

5.1 One-sample t-test

Examples:

```
# t distribution with DF=6

> qt(0.975, df=6) # t_{0.975}(6)
[1] 2.446912
> qt(0.025, df=6) # t_{0.025}(6)
[1] -2.446912

# stnadard normal distribution

> qnorm(0.975) # z_{0.975}
[1] 1.959964
> qnorm(0.025) # z_{0.025}
[1] -1.959964
```

Summary: Testing Procedure (Two-sided)

Suppose $H_0 : \mu = \mu_0$ (μ_0 is a constant) and $H_1 : \mu \neq \mu_0$, and set a ***significant level*** α (e.g., 0.01, 0.05). Then,

- Reject H_0 if and only if $|T| > t_{1-\alpha/2}(n-1)$.

Note:

- $P(t_{\alpha/2} < T < t_{1-\alpha/2}) = 1 - \alpha$.
- The idea is to reject H_0 if T falls in the top 100α % extreme range.

5.1 One-sample t-test

Summary: Testing Procedure (One-sided)

Suppose $H_0 : \mu = \mu_0$ and $H_1 : \mu < \mu_0$, and the significant level is α. Then,

- Reject H_0 if and only if $T < t_\alpha(n-1)$.

Note:

- For example, if $df = 6$ and $\alpha = 0.05$, reject if $T < t_{0.05}(6) = -1.943$.
- The idea is to reject H_0 if T (and hence $\bar{x}$) falls in the smallest 5% of the range because a small T (and $\bar{x}$) supports H_1 against H_0.
- The definition of H_0 and H_1 presumes that μ is never larger than μ_0.

The opposite case:

Suppose $H_0 : \mu = \mu_0$ and $H_1 : \mu > \mu_0$. Then,

- Reject H_0 if and only if $T > t_{1-\alpha}(n-1)$.

Sample Code 1: By ”t.test” function

```
> t.test(Income, mu = 79500)

        One Sample t-test

data:  Income
t = -0.9465, df = 6, p-value = 0.3804
alternative hypothesis: true mean is not equal to 79500
95 percent confidence interval:
 46977.71 93879.43
sample estimates:
mean of x
 70428.57
```

5.1 One-sample t-test

p-value:

- The ***p-value*** is
  - $P(|T| > |t|)$ under H_0 if $H_1 : \mu \neq \mu_0$,
  - $P(T > t)$ under H_0 if $H_1 : \mu > \mu_0$, and
  - $P(T < t)$ under H_0 if $H_1 : \mu < \mu_0$.
- The p-value ranges in $[0, 1]$.
- The p-value represents the probability that a more extreme event than T happens under H_0.
- In general, ***H_0 is rejected if and only if p-value $< \alpha$.***

5.1 One-sample t-test

Example: Math PhD salary

When $H_1 : \mu \neq 79500$,

$$\begin{aligned} \text{p-value} &= P(|T| > |-.94652|) = 2P(T > .94652) \\ &= 2(1 - P(T < 0.9452)) = 0.3804205 \end{aligned}$$

Sample code:

```
> N <- length(Income)
> DF <- N-1
> T <- (mean(Income)-79500)/(sd(Income)/sqrt(N)) # t-statistics
> Thre <- qt(0.975, df=DF) # threshold for rejecting H_0
> P.val <- 2*(1- pt(abs(T),df=DF))
> c(N,DF,T,Thre,P.val,mean(Income))
[1]     7.0000000     6.0000000    -0.9465320     2.4469119     0.3804205
[6] 70428.5714286
```

5.1 One-sample t-test

Example: Hypertension

The average systolic blood pressure of hypertension patients in a hospital is 162.6 mmHg. 10 patients in the hospital took a prescription, and their blood pressure after 4 weeks are:
172, 150, 154, 147, 168, 141, 156, 164, 146, 159.
Test if the blood pressure decreased significantly after taking the prescription.

(Solutions:)

Let μ be the population average blood pressure after prescription. Want to test $H_0 : \mu = 162.6$ against $H_1 : \mu < 162.6$.

5.1 One-sample t-test

Sample code:

```
> X <- c(172, 150, 154, 147, 168, 141, 156, 164, 146, 159)
> t.test(X, mu=162.6, alternative="less")

One Sample t-test

data:  X
t = -2.1604, df = 9, p-value = 0.02952
alternative hypothesis: true mean is less than 162.6
95 percent confidence interval:
     -Inf 161.5548
sample estimates:
mean of x
    155.7
```

H_0 is rejected with a significance level 0.05 since p-value is less than 0.05.

5.1 One-sample t-test

Confidence Interval:

We often want to construct a 95% (or in general $\mathbf{100 \times (1 - \alpha)}$%) ***confidence interval*** $\mathbf{[a, b]}$ for μ to say "μ is within $\mathbf{[a, b]}$ with 95% probability".

The result of the sample code 1 (p.14) reads that the 95% confidence interval for μ is $\mathbf{[46977.71, 93879.43]}$, meaning μ is within this range with 95% probability.

5.1 One-sample t-test

P(46977 $\leq \mu \leq$ 93879) $=$ 0.95?

No. μ is a parameter, which is a constant, and 46977 and 93879 are also constants. There are no randomness for the condition '**46977** $\leq \mu \leq$ **93879**'. $\boldsymbol{P(46977 \leq \mu \leq 93879)}$ should be either 0 or 1.

P($A \leq \mu \leq B$) $=$ 0.95:

However, 46977 and 93879 are realization of random variables. The formal definition of a confidence interval is as follows.

Definition: Confidence Interval

Suppose A and B are functions of $X_1, \cdots, X_n$, so A and B are random variables. When

$$P(A \leq \mu \leq B) = 1 - \alpha,$$

we say $[A, B]$ is a $100 \times (1 - \alpha)$% confidence interval for μ.

5.1 One-sample t-test

Note:

Different intervals $[A_1, B_1]$ and $[A_2, B_2]$ may satisfy

$$P(A_1 \le \mu \le B_1) = P(A_2 \le \mu \le B_2) = 1 - \alpha,$$

so often the shortest interval which satisfies the definition is called a confidence interval.

Example:

If $X \sim N(\mu, 1)$ (and hence $Z = X - \mu \sim N(0, 1)$), then

$$P(X - 1.96 \le \mu \le X + 1.96) = P(-1.96 \le Z \le 1.96) = 0.95.$$
$$P(X - 2.33 \le \mu \le X + 1.75) = P(-1.75 \le Z \le 2.33) = 0.95.$$

5.1 One-sample t-test

Theorem:

Suppose $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$. Then, the $\mathbf{100 \times (1 - \alpha)}$% confidence interval of μ is

$$\left[\bar{X} - t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}},\ \ \bar{X} + t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]$$

where the degrees of freedom for $t_{1-\alpha/2}$ is $n - 1$.

Note:
The normal assumption is usually hard to verify for a small sample, but the violation of the assumption does not impact the conclusion much if there are no extreme observations.

5.1 One-sample t-test

Proof of the Theorem:

$$
\begin{aligned}
& X_1, \cdots, X_n \sim N(\mu, \sigma^2) \\
\Rightarrow \quad & \bar{X} \sim N(\mu, \sigma^2/n) \\
\Rightarrow \quad & \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \qquad (2) \\
\Rightarrow \quad & P\left(-t_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{1-\alpha/2}\right) = 1 - \alpha. \\
\Rightarrow \quad & P\left(\bar{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha. \quad \blacksquare
\end{aligned}
$$

We need some algebra to show (2).

5.1 One-sample t-test

Example: Math PhD salary

```
> CI.lower <- mean(Income) - qt(0.975,df=6) * sd(Income)/sqrt(7)
> CI.upper <- mean(Income) + qt(0.975,df=6) * sd(Income)/sqrt(7)
> c(CI.lower,CI.upper)
[1] 46977.71 93879.43
```

5.2 Wilcoxon signed-rank test

When normality assumption of t-test is severely violated, ***Wilcoxon signed-rank test*** can be used.

This test only assumes symmetry of the distribution of observations (so we say this test is ***robust*** against violation of normality).

However, this test has a ***lower power*** than one-sample t-test. That is, it is harder to reject H_0 even when the data are the same.

5.2 Wilcoxon signed-rank test

Theorem:

Suppose $X_1, \cdots, X_n$ are IID and symmetric continuous random variables. We want to test $H_0 : \mu = \mu_0$ with a significant level α.

The ***signed rank*** (R_i) of X_i is the rank of $|X_i - \mu_0|$ in increasing order multiplied by 1 (if $X_i - \mu_0 \geq 0$) or -1 (if $X_i - \mu_0 < 0$).

$$
\begin{aligned}
V &:= \sum_{R_i > 0} |R_i| \\
V^* &:= \sum_{R_i < 0} |R_i|
\end{aligned}
$$

The distribution of V does not depend on the distribution of X_i's. Note that $V + V^* = \frac{n(n+1)}{2}$ is constant, so a smaller $\min(V, V^*)$ means a larger difference between V and V^* and hence a larger difference between the 2 groups.

P-values:

Let V be the statistic under H_0 (random variable), and v be the statistic from the sample (constant). The p-values are defined as follows.

- For two-sided test $H_1 : \mu \neq \mu_0$,
  the p-value is $P(\min(V, V^*) \leq v) = 2P(V \leq v)$.
- For two-sided test $H_1 : \mu < \mu_0$,
  the p-value is $P(V \leq v)$.
- For two-sided test $H_1 : \mu > \mu_0$,
  the p-value is $P(v \leq V)$.

As a consequence of this, the rejection regions are defined as follows.

Rejection region (two-sided test):

Consider an alternative hypothesis $H_1 : \mu \neq \mu_0$.

Let α be the significance level (e.g., $\alpha = 0.05$). Find the largest integer $v_{\alpha/2} > 0$ such that:

$$P(\min(V, V^*) \leq v_{\alpha/2}) \leq \alpha \quad \text{under } H_0.$$

Or equivalently,

$$P(V \leq v_{\alpha/2}) \leq \frac{\alpha}{2} \quad \text{under } H_0.$$

Reject H_0 if and only if $V \leq v_{\alpha/2}$.

Rejection region (one-sided test):

- Consider an alternative hypothesis $H_1 : \mu < \mu_0$. Find the largest integer $v_\alpha > 0$ such that:

$$P(V \le v_\alpha) \le \alpha \quad \text{under } H_0.$$

  Reject H_0 if and only if $V \le v_\alpha$.

- Consider an alternative hypothesis $H_1 : \mu > \mu_0$. Find the smallest integer $v_{1-\alpha} > 0$ such that:

$$P(v_{1-\alpha} \le V) \le \alpha \quad \text{under } H_0.$$

  Reject H_0 if and only if $v_{1-\alpha} \le V$.

5.2 Wilcoxon signed-rank test

Example: Math PhD Salary

Data: 45000, 80000, 50000, 95000, 65000, 110000, 48000.

$H_0 : \mu = 79500$.

i	X_i	$X_i - 79500$	signed rank
1	45000	-34500	-7
2	80000	500	1
3	50000	-29500	-4
4	95000	15500	3
5	65000	-14500	-2
6	110000	30500	5
7	48000	-31500	-6

$$V = 9, \quad V^* = 19, \quad \min(V, V^*) = 9.$$

Given n, a small $\min(V, V^*)$ rejects H_0 for a two-sided test.

5.2 Wilcoxon signed-rank test

When the 7 observations come from any symmetric distribution centered at 79500 (this is H_0), a simulation shows that

$$P(V \leq 2) \leq 0.025 \leq P(V \leq 3) \leq 0.05 \leq P(V \leq 4)$$

Consequently, $P(V^* \leq 2) \leq 0.025$ and $P(\min(V, V^*) \leq 2) \leq 0.05$.

Therefore,

- the two-sided test rejects H_0 if and only if $\min(V, V^*) \leq 2$ (that is, $V \leq 2$ or $V \geq 26$).
- a one-sided test with $H_0 : \mu < \mu_0$ rejects H_0 if and only if $V \leq 3$.

5.2 Wilcoxon signed-rank test

Sample Code: Two-sided test ($H_1 : \mu \neq 79500$)

```
> wilcox.test(Income, mu=79500)

        Wilcoxon signed rank test

data:  Income
V = 9, p-value = 0.4688
alternative hypothesis: true location is not equal to 79500
```

Note:

- The output indicates $P(\min(V, V^*) \leq 9) = 2P(V \leq 9) = 0.4688$, so H_0 is not rejected.
- The p-value is larger than the p-value of the t-test. This is consistent with a weaker power of the Wilcoxon test.

Sample Code: One-sided test ($H_1 : \mu < 79500$)

```
> wilcox.test(X, mu=79500, alternative="less")

Wilcoxon signed rank test

data:  X
V = 9, p-value = 0.2344
alternative hypothesis: true location is less than 79500
```

Note:

- The output indicates $P(V \leq 9) = 0.2344$, so H_0 is not rejected.
- The p-value is just a half of the p-value by the two-sided test, because of an additional assumption that $\mu > 79500$ never happens.

5.2 Wilcoxon signed-rank test

Sample Code: One-sided test (H_1 : $\mu > 79500$)

```
> wilcox.test(X, mu=79500, alternative="greater")

Wilcoxon signed rank test

data:  X
V = 9, p-value = 0.8125
alternative hypothesis: true location is greater than 79500
```

Note:

- The output indicates $P(9 \leq V) = 0.8125$, so H_0 is not rejected.
- $V < V^*$ is an indication that $\mu < 79500$. Therefore, $H_1 : \mu > 79500$ is not supported at all, making the p-value very large.

Exercise 1:

Do a simulaiton to calculate $P(V \leq 9)$ when the sample size is 7. I.e., repeat, say 10,000 times, a procedure to generate 7 random numbers from a symmetric distribution with mean zero and calculate V.

Exercise 2:

Use a different symmetric distribution with mean zero to do the same simulation as Exercise 1. The results are similar or much different?

5.2 Wilcoxon signed-rank test

Note:
Wilcoxon signed rank test assumes there are no ties ('ties' mean observations with the same absolute deviation or observation(s) which exactly equal to the population mean).

If the data have ties, we should use an adjusted version of the test: ***'wilcox.exact'*** in the ***'exactRankTests'*** package.

5.2 Wilcoxon signed-rank test

Review:

Most ***hypothesis testing*** has the following procedure:

1. Calculate a ***test statistic*** .
2. Determine the ***distribution of the test statistic*** under $\boldsymbol{H_0}$, either mathematically or by simulation.
3. ***Reject*** $\boldsymbol{H_0}$ if the test statistic is in an ***extreme range*** in favor of $\boldsymbol{H_1}$ (i.e., 5% most extreme range if $\alpha = \mathbf{0.05}$).

5.3 Two-sample test

Suppose we want to test the hypothesis $\mu_X - \mu_Y = d$ (d is a constant, usually zero) with the following two-sample data.

	Group 1	Group 2
Sample	$x_1, \cdots, x_n$	$y_1, \cdots, y_m$
Sample mean	$\bar{x}$	$\bar{y}$
Sample S.D.	s_x	s_y
Population dist'n	$N(\mu_X, \sigma_X^2)$	$N(\mu_Y, \sigma_Y^2)$

5.3 Two-sample test

Case 1: $\sigma_X = \sigma_Y$

If we have a strong belief that $\sigma = \sigma_X = \sigma_Y$, a good estimate for σ^2 is:

$$
\begin{aligned}
s^2 &= \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \\
&= \frac{1}{n+m-2}\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{j=1}^{m}(y_j - \bar{y})^2\right\}
\end{aligned}
$$

Then we may use the following fact.

Theorem:

If all X_i's and Y_j's are independent and $\sigma_X = \sigma_Y$,

$$T := \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$$

where $t(n + m - 2)$ means t-distribution with degrees of freedom $n + m - 2$.

Note: $H_0 : \mu_X = \mu_Y$ is rejected against $H_1 : \mu_X \neq \mu_Y$ if $|T| > t_{1-\alpha/2}$.

Example:

The following data are the average math test scores for 5-th graders in 7 Detroit and 5 Dearborn schools. Test if the two groups have significant difference with a significance level at 5%.

Detroit: 519.7; 494.7; 501.5; 499.1; 493.1; 505.0; 529.1.
Dearborn: 530.1; 507.6; 525.0; 518.1; 520:4.

5.3 Two-sample test

Sample Code:

```
> DT <- c(519.7, 494.7, 501.5, 499.1, 493.1, 505.0, 529.1)
> DB <- c(530.1, 507.6, 525.0, 518.1, 520.4)
> t.test(DT, DB,var.equal=T)

        Two Sample t-test

data:  DT and DB
t = -2.0747, df = 10, p-value = 0.06477
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -29.473808   1.050951
sample estimates:
mean of x mean of y
 506.0286  520.2400
```

The difference is not significantly different at 5% significance level since p-value $> \mathbf{0.05}$.

5.3 Two-sample test

Exercise 3:

Calculate a ***T***-statistic of a two-sample test for the above example, by only using ***mean*** and ***sd*** functions as well as basic mathematical functions (square root, $+$, $-$, $*$, $/$).

Then, calculate the $\boldsymbol{p - value}$ by using the ***pt*** function.

5.3 Two-sample test

Case 2: $\sigma_X \neq \sigma_Y$

If $\sigma_X \neq \sigma_Y$, there is not beautiful mathematical theory for testing $\mu_X = \mu_Y$, but the following approximate result can be used.

Theorem:

$$T := \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{s_Y^2}{m}}} \sim t(k)$$

where

$$k = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}.$$

The t-test based on the above result is called ***"Welch's t-test"***.

5.3 Two-sample test

Note: The t-distribution with non-integer degrees of freedom k is defined by its density function:

$$f_k(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}\left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

Example:

```
> t.test(DT, DB) # Welch's t-test

        Welch Two Sample t-test

data:  DT and DB
t = -2.2462, df = 9.917, p-value = 0.0487
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -28.32447136  -0.09838578
sample estimates:
mean of x mean of y
 506.0286  520.2400
```

$H_0 : \mu_X = \mu_Y$ is rejected.

5.4 Comparison of variances

Under the same setting as Section 5.3, we may want to test $H_0 : \sigma_X^2 = \sigma_Y^2$ against $H_1 : \sigma_X^2 \neq \sigma_Y^2$. We can use the following fact.

Theorem:
Under the setting in Section 5.3,

$$F := \frac{S_X^2}{S_Y^2} \sim F(n-1, m-1). \tag{3}$$

where $F(n-1, m-1)$ is an F-distriubtion with the degrees of freedom $n-1$ (numerator) and $m-1$ (denominator).

5.4 Comparison of variances

The P-value:

Suppose that F is the statistic under H_0 and f is the statistic from the sample.

$$2 \min \{P(F < f), P(f < F)\}$$

Rejection regions:

We reject H_0 against H_1 at significance level α (e.g., 0.05) if and only if

$$F < F_{\alpha/2}(n-1, m-1) \quad \text{or} \quad F > F_{1-\alpha/2}(n-1, m-1)$$

Note: One-sided tests are also possible.

5.4 Comparison of variances

Example: Math score in Detroit and Dearborn

(Sample code:)

```
> var.test(DT,DB) # two-sided

        F test to compare two variances

data:  DT and DB
F = 2.5429, num df = 6, denom df = 4, p-value = 0.3855
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.2764875 15.8353062
sample estimates:
ratio of variances
          2.542941
```

H_0 is not rejected since p-value $> \mathbf{0.05}$.

5.4 Comparison of variances

(Sample code (continued):)

```
> N <- length(DT)
> M <- length(DB)
> var(DT)/var(DB)
[1] 2.542941

> qf(0.975, df1=N-1, df2=M-1) # thresholds for rejection
[1] 9.197311
> qf(0.025, df1=N-1, df2=M-1) # thresholds for rejection
[1] 0.1605868

> 2*(1- pf(2.542941, df1=N-1, df2=M-1)) # p-value
[1] 0.3855074
```

The rejection region for this test is:

$$F < 0.1605, \quad \text{or} \quad F > 9.1973$$

5.5 Two-sample Wilcoxon test

If the normal assumption in 5.3 is not satisfied, you can use the ***Wilcoxon rank-sum test*** (also called ***Mann-Whitney test***).

We keep the notation μ_X and μ_Y for location parameters. Those can be the population means, but may be the population medians for example. Since a normal distribution is not assumed, the population means may not exist.

Assumptions:

- Data are paired and come from the same population.
- Pairs are independent.

The assumption does not require a symmteric distribution for observations.

Procedure:
Combine all observations X_i's and Y_j's (let $N := n + m$ and rank all observations by increasing order. Let

$$
\begin{aligned}
W &:= \sum_{i=1}^{n} rank(X_i) - \frac{n(n+1)}{2} \\
W^* &:= \sum_{j=1}^{m} rank(Y_j) - \frac{m(m+1)}{2}
\end{aligned}
$$

Our H_0 is that 'X and Y have the same location parameter (such as population mean)', and H_1 is that 'X and Y have significantly different location parameters' if the test is two-sided.

5.5 Wilcoxon signed-rank test

P-values:

Let W be the statistic under H_0 (random variable), and w be the statistic from the sample (constant) . The p-values are defined as follows.

- For two-sided test $H_1 : \mu \neq \mu_0$,
  the p-value is $P(\min(W, W^*) \leq w) = 2P(W \leq w)$.
- For two-sided test $H_1 : \mu < \mu_0$,
  the p-value is $P(W \leq w)$.
- For two-sided test $H_1 : \mu > \mu_0$,
  the p-value is $P(w \leq W)$.

As a consequence, the rejection regions are defined as follows.

5.5 Two-sample Wilcoxon test

Rejection region (two-sided test)

- For $H_1 : \mu_X \neq \mu_Y$:
  - The rejection region is $W \leq w_{\alpha/2}$ where $w_{\alpha/2}$ is the largest integer such that, under H_0,

$$P(W \leq w_{\alpha/2}) \leq \frac{\alpha}{2} \quad \text{or equivalently} P(\min(W, W^*) \leq w_{\alpha/2}) \leq \alpha.$$

- For $H_1 : \mu_X < \mu_Y$:
  - The rejection region is $W \leq w_\alpha$ where w_α is the largest integer such that, under H_0,

$$P(W \leq w_\alpha) \leq \alpha.$$

- For $H_1 : \mu_X > \mu_Y$, the rejection regi
  - The rejection region is $W \leq w_{1-\alpha}$ where $w_{1-\alpha}$ is the largest integer such that, under H_0,

$$P(w_{1-\alpha} \leq W) \leq \alpha.$$

w_α can be found by simulation.

5.5 Two-sample Wilcoxon test

Example: Math score in Detroit and Dearborn

```
> DT
[1] 519.7 494.7 501.5 499.1 493.1 505.0 529.1
> DB
[1] 530.1 507.6 525.0 518.1 520.4
> rank(c(DT,DB))
 [1]  8  2  4  3  1  5 11 12  6 10  7  9
> sum(rank(c(DT,DB))[1:7]) -  7*(7+1)/2
[1] 6
> sum(rank(c(DT,DB))[8:12]) - 5*(5+1)/2
[1] 29
> wilcox.test(DT,DB) # two-sided Wilcoxon rank-sum

        Wilcoxon rank sum test

data:  DT and DB
W = 6, p-value = 0.07323
alternative hypothesis: true location shift is not equal to 0
```

Example: Math score in Detroit and Dearborn (Continued)

```
> wilcox.test(DT,DB,alternative="less") # H_1: mu_X < mu_Y

Wilcoxon rank sum test

data:  DT and DB
W = 6, p-value = 0.03662
alternative hypothesis: true location shift is less than 0

> wilcox.test(DT,DB,alternative="greater") # H_1: mu_X > mu_Y

Wilcoxon rank sum test

data:  DT and DB
W = 6, p-value = 0.976
alternative hypothesis: true location shift is greater than 0
```

Exercise 4:

Suppose $n = 7$ and $m = 5$. Calculate $w_{0.01}$, $w_{0.05}$ and $w_{0.10}$ by simulation.

5.6 The paired t test

IF two samples are two variables from the same experimental units, you can take the difference between two samples then do a one-sample test.

Example: Average Wear for Two Types of Tires

Automobile	Tire A	Tire B	$D = A - B$
1	10.6	10.2	0.4
2	9.8	9.4	0.4
3	12.3	11.8	0.5
4	9.7	9.1	0.6
5	8.8	8.3	0.5
mean	$\bar{x}_1 = 10.24$	$\bar{x}_2 = 9.76$	$\bar{d} = 0.48$
S.D.	$s_1 = 1.316$	$s_2 = 1.328$	$s = 0.083$

5.6 The paired t test

```
> A <- c(10.6, 9.8, 12.3, 9.7, 8.8)
> B <- c(10.2, 9.4, 11.8, 9.1, 8.3)
> t.test(A-B)

        One Sample t-test

data:  A - B
t = 12.8285, df = 4, p-value = 0.0002128
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.3761149 0.5838851
sample estimates:
mean of x
     0.48
```

Note: “t.test(A, B, paired=TRUE)” makes the same result.

5.7 The matched-pairs Wilcoxon test

For a paired samples whose difference do not follow a normal distribution, we can use Wilcoxon signed rank test.

Recall that while normality is not required, it is still required that the distribution is symmetric.