

chapter 7

Hypothesis testing using R

Instructor: Li, Han

Contents

- ❑ Null hypothesis and alternative hypothesis
- ❑ Steps for hypothesis testing
- ❑ One/Two sample t test
- ❑ Paired t test

Hypothesis testing

A statistical hypothesis is a hypothesis that is testable by observing the data. A statistical hypothesis test is a method of statistical inference.

Commonly, we propose two contradictive statistical hypothesis and use the a data set to compare which hypotheis is more likely to produce the observed data.

A research hypothesis typically states that there is a real change, a real difference, or real effect in the population or process. The hypothesis is often called H_1 .

The opposite, null hypothesis, called H_0 , then states that there is no real change, difference, or effect.

The basic strategy of hypothesis testing is to try to support a research hypothesis by showing that the sample results are highly unlikely, given the null hypothesis is true. However, the results are more likely to happen, if the research hypothesis is true. In a word, we try to reject H_0 .

Steps for hypothesis testing

- ❑ Declare the hypothesis
- ❑ Select a test statistic
- ❑ Specify the confidence level α
- ❑ Specify the rejection region of the test statistic
- ❑ Given the data, make a conclusion

Steps for hypothesis testing

- Step 1. Declare the null hypothesis H_0 and the alternative hypothesis H_1 .

In a mean estimation problem, for example the average housing price per square meters in Shenzhen in November, 2017, we may consider $H_0: \mu \leq 35,000$ and $H_1: \mu > 35,000$.

A hypothesis that completely specifies the parameters is called **simple**. If it leaves some parameter undetermined it is **composite**.

A hypothesis is **one-sided** if it proposes that a parameter is larger than or less than a given value; it is **two-sided** if it says the parameter is not equal to a given value.

This difference results in different methods to calculate their p-values.

- Step 2. Select a test statistic.

This is a summary statistic calculated from the data and it helps us to draw the conclusion whether to reject the null hypothesis or not. For the mean estimation problem, the choice would be the sample mean.

Rejecting H_0 when it is actually true is called a **Type I Error**. It means that we decide "there is some change" (accept H_1) but no change does happen (H_0 is true).

- ▣ Step 3. Specify an acceptable level of Type I error, α , normally, 0.05 or 0.01.

α is the threshold used in deciding to reject H_0 or not. It is also called the confidence/significance level. Given $\alpha = 0.05$, if the probability of obtaining our data or more extreme data when H_0 is true, is 0.0001 (that's p-value), we reject H_0 . If p-value $> \alpha$, we think that the observed data is still normal, thus H_0 is probably to be true and we will not reject it.

-
- Step 4. Identify the value of the test statistic that lead to rejection of the null hypothesis.

For the mean estimation problem, we reject the null hypothesis if the sample mean is too large such that its p-value is less than α . The starting value to reject H_0 is called the critical value.

p-value

p-value = $P(\text{test statistic equals the sample value or is more extreme} \mid H_0 \text{ is true})$

If $p\text{-value} < \alpha$, we reject the null hypothesis, otherwise we retain H_0 . p-value can be regarded as a measure of strength to support H_0 .

-
- Step 5. Obtain the data, calculate the value of the test statistic and compare it with the critical value or find its p-value. Finally draw the conclusion whether reject H_0 or not.

t.test()

The R function `t.test()` can be used to perform both one and two sample t-tests.

The function contains a variety of options and can be called as follows:

```
t.test(x, y = NULL, alternative = c("two.sided", "less",  
"greater"), mu = 0, paired = FALSE, var.equal = FALSE,  
conf.level = 0.95)
```

```
t.test(x, y = NULL, alternative = c("two.sided", "less",  
"greater"), mu = 0, paired = FALSE, var.equal = FALSE,  
conf.level = 0.95)
```

Here **x**, **y** are numeric vectors. If **y** is excluded, the function performs a one-sample t-test only using **x**. If **y** is included, it performs a two-sample t-test using both **x** and **y**.

```
t.test(x, y = NULL, alternative = c("two.sided", "less",  
"greater"), mu = 0, paired = FALSE, var.equal = FALSE,  
conf.level = 0.95)
```

The option **mu** is the value to compare. The option **alternative** specifies the alternative hypothesis, and must be one of the following: "two.sided" (which is the default), "greater" or "less". For example, `t.test(x, alternative = "less", mu = 10)`. It means

$H_0: \mu \geq 10$, $H_1: \mu < 10$.

```
t.test(x, y = NULL, alternative = c("two.sided", "less",  
"greater"), mu = 0, paired = FALSE, var.equal = FALSE,  
conf.level = 0.95)
```

The option **paired** indicates whether or not we want a paired t-test (TRUE = yes and FALSE = no). Its default value is FALSE. The option **var.equal** indicates whether we assume the variance of x, y are equal or not.

```
t.test(x, y = NULL, alternative = c("two.sided", "less",  
"greater"), mu = 0, paired = FALSE, var.equal = FALSE,  
conf.level = 0.95)
```

Finally, the option `conf.level=1- α` determines the confidence level of the hypothesis testing and the reported confidence interval for μ in the one-sample case and $\mu_1 - \mu_2$ in the two-sample case.

One sample t test

Assume we have n observations denoted as x_1, x_2, \dots, x_n .
Suppose

$$x_i \stackrel{iid}{\sim} N(\mu, \sigma^2), i = 1, 2, \dots, n.$$

We can derive that

$$\frac{\bar{x} - \mu}{\sqrt{\sigma^2 / n}} \sim N(0,1), \quad \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \sim \chi^2(n-1),$$

Thus, by definition,

$$\begin{aligned} \frac{(\bar{x} - \mu) / \sqrt{\sigma^2 / n}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / [\sigma^2 (n-1)]}} &= \frac{\bar{x} - \mu}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1) \right] / n}} \\ &= \frac{\bar{x} - \mu}{S / \sqrt{n}} \sim T(n-1) \end{aligned}$$

Example 1. (One sample t test)

An outbreak of Salmonella-related illness was attributed to ice cream produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches of ice cream. The levels (in MPN/g) were:

*0.593 0.142 0.329 0.691 0.231 0.793 0.519
0.392 0.418*

Is there evidence that the mean level of Salmonella in the ice cream is greater than 0.3 MPN/g?

Let μ be the mean level of Salmonella in all batches of ice cream. Here the hypothesis of interest can be expressed as:

$$H_0: \mu \leq 0.3, H_1: \mu > 0.3.$$

Example 1

```
x = c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519,  
0.392, 0.418)
```

```
t.test(x, alternative="greater", mu=0.3)
```



RGui (32-bit) - [R Console]



File Edit View Misc Packages Windows Help



```
> x = c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418)
> t.test(x, alternative="greater", mu=0.3)
```

One Sample t-test

```
data: x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133      Inf
sample estimates:
mean of x
0.4564444
```

```
> |
```

From the output we see that the p-value = 0.029. Hence, there is moderately strong evidence that the mean Salmonella level in the ice cream is above 0.3 MPN/g.

Two sample pooled t test

Assume for data set 1, we have n observations denoted as x_1, x_2, \dots, x_n ; for data set 2, we have m observations denoted as y_1, y_2, \dots, y_m . Suppose

$$x_i \stackrel{iid}{\sim} N(\mu_1, \sigma_1^2), i = 1, 2, \dots, n;$$

$$y_i \stackrel{iid}{\sim} N(\mu_2, \sigma_2^2), i = 1, 2, \dots, m.$$

Assuming that $\sigma_1^2 = \sigma_2^2$, we can derive that

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(n + m - 2),$$

$$\text{where } s = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n + m - 2}.$$

Example 2. (Two sample t test)

We want to check the effect of a new drug. 6 subjects were given the drug (treatment group) and an additional 6 subjects a placebo (control group). Their reaction time to a stimulus was measured (in ms). We want to perform a two-sample t-test for comparing the means of the treatment and control groups.

Let μ_1 be the mean of the population taking medicine and μ_2 the mean of the untreated population. Here the hypothesis of interest can be expressed as:

$$H_0: \mu_1 - \mu_2 \geq 0, H_1: \mu_1 - \mu_2 < 0.$$

Example 2.1

Treat = c(91, 87, 99, 77, 88, 91)

Control= c(101, 110, 103, 93, 99, 104)

t.test(Treat, Control, alternative="less",
var.equal=TRUE)

In this example, we assume the variances of x and y are equal, thus $df=n+m-2$.



RGui (32-bit) - [R Console]



File Edit View Misc Packages Windows Help



```
> Treat = c(91, 87, 99, 77, 88, 91)
> Control= c(101, 110, 103, 93, 99, 104)
> t.test(Treat, Control, alternative="less", var.equal=TRUE)
```

Two Sample t-test

data: Treat and Control

t = -3.4456, df = 10, p-value = 0.003136

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -6.082744

sample estimates:

mean of x mean of y

88.83333 101.66667

```
> |
```

Welch t test

If the variance of the two samples are not equal, we need to change the formula for calculating t statistic and df. In this case,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \sim T(df),$$

$$\text{where } df = \frac{(s_x^2 / n + s_y^2 / m)^2}{(s_x^2 / n)^2 / (n - 1) + (s_y^2 / m)^2 / (m - 1)}.$$

To check whether the variance of the two samples are equal, we can use the box plot of the data or the F test.

Example 2.2

```
Treat = c(91, 87, 99, 77, 88, 91)
Control= c(101, 110, 103, 93, 99, 104)
t.test(Treat, Control, alternative="less")
```

In this example, we donot assume the variances of x and y are equal, thus the df changes!



RGui (32-bit) - [R Console]



File Edit View Misc Packages Windows Help



```
> t.test(Treat, Control, alternative="less")
```

```
Welch Two Sample t-test
```

```
data: Treat and Control
```

```
t = -3.4456, df = 9.4797, p-value = 0.003391
```

```
alternative hypothesis: true difference in means is less than 0
```

```
95 percent confidence interval:
```

```
-Inf -6.044949
```

```
sample estimates:
```

```
mean of x mean of y
```

```
88.83333 101.66667
```

```
> |
```

Here the pooled t-test and the Welsh t-test give roughly the same results (p-value = 0.0031 and 0.0033, respectively).

Paired t test

In a survey, one can measure each subject twice, before and after a treatment. In either of these situations we cannot use two-sample t-test since the independence assumption is not valid. Instead we need to use a paired t-test. This can be done using the option `paired = TRUE`. For paired t test, we first calculate the difference of the same object under two conditions, then use the one sample t test.

Example 3. (Paired t test)

A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. We use a paired t test to determine whether cars get significantly better mileage with premium gas.

Example 3

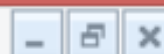
```
reg = c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
prem = c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
t.test(prem, reg, alternative="greater",
paired=TRUE)
```



RGui (32-bit) - [R Console]



File Edit View Misc Packages Windows Help



```
> reg = c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> prem = c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> t.test(prem, reg, alternative="greater", paired=TRUE)
```

Paired t-test

```
data:  prem and reg
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.180207      Inf
sample estimates:
mean of the differences
                2
```

```
> |
```

The results show that the t-statistic is equal to 4.47 and the p-value is 0.00077. Since the p-value is very low, we reject the null hypothesis. There is strong evidence of a mean increase in gas mileage between regular and premium gasoline.

Summary

In this session, we have learned

- ▣ the procedures of hypothesis testing
- ▣ how to conduct one/two/paired sample t test.