

Hypothesis Testing using R

Pramudita Satria Palar, Vani Viridyawan, Ferryanto

2022-04-24

Introduction

In this tutorial, we will be testing a hypothesis using R. There are several hypothesis testing methods that you can immediately use once you have R installed on your computer. You can use these functions for various applications according to your needs (e.g., one sample or two sample difference test or normality test). First, let's return to the fundamental question: Why do we need hypothesis testing? Recall that the sample we take from the population is limited in number, so we need to be careful in drawing conclusions from our data. Hypothesis testing helps in answering these questions.

It should be noted that the question or hypothesis that you want to answer must be able to be answered using a scientific method, for example: "Does students who take the Calculus course tutorial get a higher score than those who do not take the tutorial?" (in a statistical sense). The output of the hypothesis test is whether the difference that occurs is statistically significant or not. These results will help us in making decisions later.

Defining hypotheses

First, we need to determine the **null** and **alternative** hypotheses in testing the hypothesis. Usually, though not always, the null hypothesis is something that we want to reject. For example, if you create a manufacturing method that is better than the old method, you make the null hypothesis that the new and old manufacturing methods perform the same; therefore, you want to reject the null hypothesis. An alternative hypothesis is a hypothesis that you accept when you reject the null hypothesis because there is sufficient evidence for you to do so. We will use the notation \mathcal{H}_0 for the null hypothesis and \mathcal{H}_a for the alternative hypothesis.

The following examples give you an idea of \mathcal{H}_0 and \mathcal{H}_a :

- \mathcal{H}_0 : There is no significant difference between the results yielded by method A and method B.
- \mathcal{H}_a : Method A yields higher output than method B.

Notice the word *higher*, which indicates that the hypothesis test above is one-sided. Similarly, you can also define a two-sided hypothesis test, for example:

- \mathcal{H}_0 : Your monthly average wage is 1000\$.
- \mathcal{H}_a : Your monthly average wage is not 1000\$.

The alternative hypothesis stated above is two-sided because your salary could be less or higher than 1000\$.

The two possible outputs from hypothesis testing are as follows:

- A small p -value indicates strong evidence to reject H_0 , so you reject H_0 .
- A large p -value indicates weak evidence to reject H_0 , so you fail to reject H_0 .

The threshold of p -value where you decide between rejecting or failing to reject H_0 will depend on the level of significance that you want. Generally, the value 0.05 is commonly used in various communities. However, some communities take even lower limits to be more confident in their decisions.

To give an initial illustration, we will try to compare two data sets from $\mathcal{N}(0, 1)$ and $\mathcal{N}(2, 1.5)$, each with 50 observations, and we will name them X_1 and X_2 .

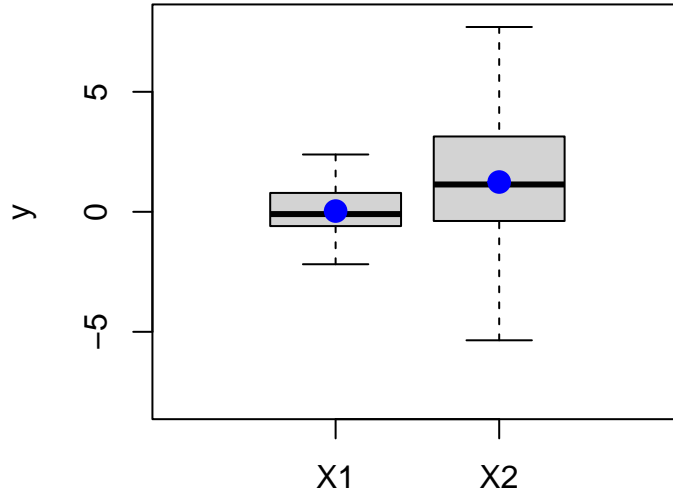


Figure 1: Boxplot of the two data that we will compare (blue filled circle shows the mean values)

The boxplot above shows that X_2 has a higher mean than X_1 , so we can say that X_2 is higher on average than X_1 . But is the difference just a coincidence? Or because a certain effect causes X_2 to be higher on average than X_1 ? Pay attention that some realizations from X_2 have lower values than all realizations from X_1 . This is an example of a two-sample test in which we want to compare whether there is a significant difference between the two samples or not. In this regard, the quantity that we want to compare is the difference between the means of the data. A hypothesis test can also be a one-sample test, which we use to investigate whether a mean value from one sample is consistent with a hypothesized value of the mean.

We will answer these questions with a t-test using the `t.test()` function.

Using the function `t.test()`

t-test with one sample

One-sample t-test usually tests one of the following pairs of null and alternative hypotheses

1. $\mathcal{H}_0 : \mu = \mu_0, \mathcal{H}_a : \mu \neq \mu_0$
2. $\mathcal{H}_0 : \mu \leq \mu_0, \mathcal{H}_a : \mu < \mu_0$
3. $\mathcal{H}_0 : \mu \leq \mu_0, \mathcal{H}_a : \mu > \mu_0$

where μ_0 is the value that we define for the null hypothesis.

We will start with a simple example by creating synthetic data generated from the normal distribution $\mathcal{N}(7, 2)$ with 50 observations. In the context of the real world, of course, all you have are the results of your observations. What you want to answer here is, for example, is it true that the true mean of this data is 7? Or we can define our null hypothesis as follows; for example, is it true that the actual mean is less than 7.2? Or is it true that the actual mean is greater than 7.2? We can write these three null hypotheses along with their alternative hypotheses in the following format:

1. $\mathcal{H}_0 : \mu = 7, \mathcal{H}_a : \mu \neq 7$
2. $\mathcal{H}_0 : \mu \leq 6.8, \mathcal{H}_a : \mu > 6.8$
3. $\mathcal{H}_0 : \mu \geq 7.2, \mathcal{H}_a : \mu < 7.2$

The figure below shows a visualization of the data that we want to analyze in the form of a histogram. At first glance, the data seems to have a mean number of $\mu = 7$. Hypothesis testing will provide a formal answer to this question.

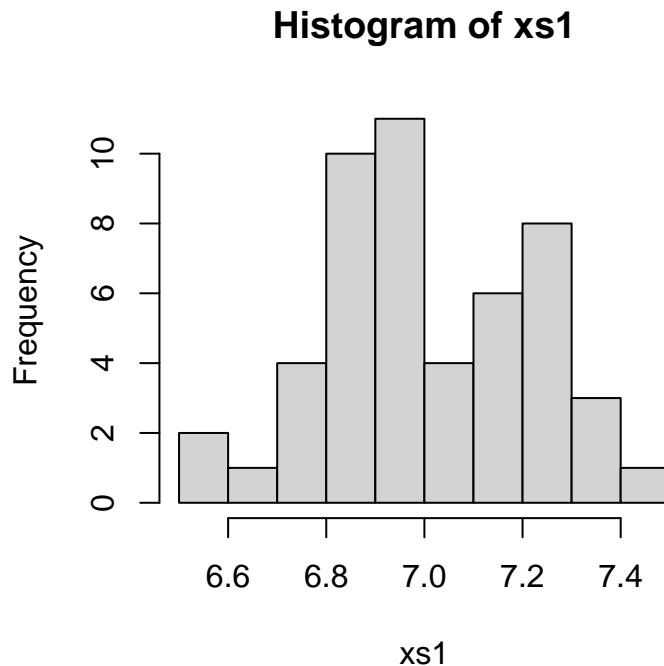


Figure 2: Boxplot that visualizes the distribution of the data used in the one-sample t-test

First, we will try using the following null hypothesis $\mathcal{H}_a : \mu = 7$, and then use the `t.test()` function. To perform a t-test with `t.test()`, it is recommended that you enter your minimum argument, i.e., `mu` (the null hypothesis) and `alternative` (that you can enter with `two.sided`, `less`, or `greater`). For example, if your null hypothesis is $\mathcal{H}_0 : \mu = 7$, you must enter `two.sided` in `alternative`. The results of the t-test will be stored in a variable which we will name `tresult`.

```
set.seed(5)
xs1 <- rnorm(50, mean=7, sd = 0.2)
tresult <- t.test(xs1, mu = 7, alternative="two.sided")
```

You can output the results of the t-test using the following syntax (or simply, just type `tresult` in your R console)

```
tresult

##
## One Sample t-test
##
## data: xs1
## t = 0.42965, df = 49, p-value = 0.6693
## alternative hypothesis: true mean is not equal to 7
```

```
## 95 percent confidence interval:
## 6.952244 7.073730
## sample estimates:
## mean of x
## 7.012987
```

Pay attention to significant figures such as p-value and t. You need to look at the p-value and compare it to the **significance level** that you already set. For example, if you set the significance level to 5 percent, you will fail to reject the null hypothesis if the p-value obtained is less than 0.05. On the other hand, you will reject the null hypothesis if the p-value is greater than 0.05. For the above case, you will get a p-value greater than 0.05, which means you failed to reject the hypothesis that $\mu = 7$. In other words, there is strong evidence for rejecting $\mu = 7$.

If you want to display the p-value and t in your console (e.g., for further calculations), you can type `tresult$p.value` and `tresult$statistic` as shown in the example below.

```
tresult$statistic
```

```
##          t
## 0.4296464
```

```
tresult$p.value
```

```
## [1] 0.6693364
```

Now, let's try with other pairs of null and alternative hypothesis. For the pair of hypotheses $\mathcal{H}_0 : \mu \leq 6.8$, $\mathcal{H}_a : \mu > 6.8$, then you must enter `alternative="greater"` in your console:

```
tresult_greater <- t.test(xs1,mu=6.8,alternative = "greater")
tresult_greater
```

```
##
## One Sample t-test
##
## data:  xs1
## t = 7.0463, df = 49, p-value = 2.812e-09
## alternative hypothesis: true mean is greater than 6.8
## 95 percent confidence interval:
## 6.96231      Inf
## sample estimates:
## mean of x
## 7.012987
```

You will get a very small p-value. For example, with a significance level of 5%, you reject the null hypothesis that $\mathcal{H}_0 : \mu \leq 6.8$ and accept the alternative hypothesis $\mathcal{H}_a : \mu > 6.8$. You can do the same for $\mathcal{H}_0 : \mu \geq 7.2$, $\mathcal{H}_a : \mu < 7.2$. Please try using the alternative `alternative="less"` in your R console.

```
tresult_less <- t.test(xs1,mu=7.2,alternative = "less")
tresult_less
```

```
##
## One Sample t-test
##
## data:  xs1
## t = -6.187, df = 49, p-value = 6.023e-08
## alternative hypothesis: true mean is less than 7.2
## 95 percent confidence interval:
##      -Inf 7.063664
## sample estimates:
```

```
## mean of x
## 7.012987
```

Two-sample problem

Now we return to the two-sample question that we wanted to answer above. We will use `t.test()` for two samples problem with the following syntax: `t.test(x,y)` where `x` dan `y` are the two samples we want to test. The plot below shows the histogram, normalized so that we have **density** in the ordinate, so as to visualize the difference between the two data better.

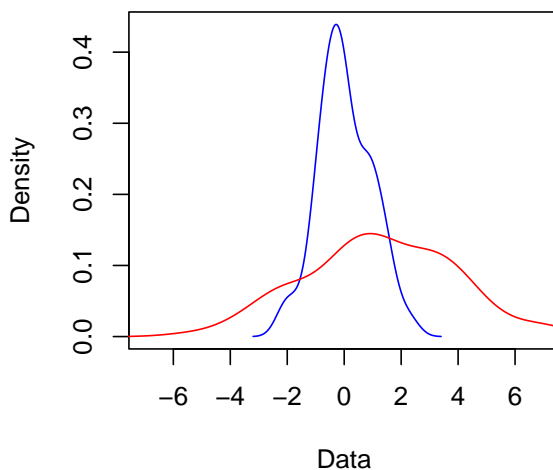


Figure 3: Density plot for the data in the two-sample problem

In this regard, we want to test the difference between two means with the following null hypothesis:

$$\mathcal{H}_0 : \mu_1 = \mu_2$$

We first try using a confidence level of 0.95 by executing the code below:

```
set.seed(5)
x1 = rnorm(100, mean = 0, sd = 1) # Data pertama
x2 = rnorm(100, mean = 1.2, sd = 2.5) # Data kedua

t2result <- t.test(x1,x2,conf.level=0.95)
```

The most important value for the results of the t-test above is the p-value, as it informs us for making a decision. The most common limit for the significance level is $\alpha = 0.05$, i.e., we reject the null hypothesis when $p < 0.05$. The initial confidence level number set by `t.test()` is 0.95, but you can adjust the `conf.level` number to your liking (e.g., `conf.level=0.99` to give a confidence level of 0.99).

Let us continue with more realistic data. The context of the following problem is to compare two catalysts to see how they affect a specific chemical process. Catalyst 1 (let us name it `cat_1`) is a frequently used catalyst. On the other hand, Catalyst 2 (let us call it `cat_2`) is cheaper and recommended for use if its performance is roughly the same as that of Catalyst 1. The experiment is performed eight times to investigate whether the second catalyst is actually better than the first catalyst. The data we want to analyze is shown in the Table below, where `num` is the number of observations.

num	cat_1	cat_2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75

We can write the observation results from the two types of catalysts using the function `c` to save the data in a vector format. The same function, i.e., `t.test()` will be used with little tweaks because we assume that the two catalysts have the same value of unknown variance. We will add the argument `var.equal=TRUE` to the function `t.test()`. Remember that `mu=0` in the context of a two-sample test is for the difference between two means

```
cat_1 <- c(91.50, 94.18, 92.18, 95.39, 91.79, 89.07, 94.72, 89.21) # Katalis 1
cat_2 <- c(89.19, 90.95, 90.46, 93.21, 97.19, 97.04, 91.07, 92.75) # katalis 2

t2resultc <- t.test(cat_1,cat_2,mu=0,var.equal=TRUE)
```

Let us show the result from the hypothesis test by typing `t2resultc` in our console:

```
t2resultc

##
## Two Sample t-test
##
## data: cat_1 and cat_2
## t = -0.35359, df = 14, p-value = 0.7289
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.373886 2.418886
## sample estimates:
## mean of x mean of y
## 92.2550 92.7325
```

We can clearly see that the p-value is higher than 0.05, which means that we fail to reject the null hypothesis. It is worth noting that our null hypothesis is $\mu_1 = \mu_2$; in other words, $\mu_1 - \mu_2 = 0$. We can then infer that the first and the second catalysts have the same performance. We can now say that the second catalyst can be used as a cheaper alternative because the hypothesis test reveals that the difference (in terms of performance) is not statistically significant.