

Class notes 2

Hypothesis testing

Introduction

One important purpose of statistical analysis is to draw inference (or conclusions) about real world phenomena using hypothesis testing. It is necessary that the phenomenon of interest can be formulated (or described) in terms of some population parameters, which in turn allows us to test hypotheses regarding their values. We limit the presentation to consider tests of a single parameter.

Concepts and terminology

Before getting started with the actual testing procedures, we need to introduce some basic concepts and terminology.

- *The null hypothesis:* The null hypothesis is a statement to be evaluated. The null hypothesis is typically a hypothesis of status quo. Statements such as “no effect” or “no difference” are common. The null hypothesis is denoted H_0 .
- *Alternative hypothesis:* The alternative hypothesis is stated as a hypothetical claim that contradicts the null hypothesis. Thus, the formulation is such that when the alternative hypothesis is true, the null hypothesis is false. The alternative hypothesis is typically denoted H_1 or H_A .

When performing a hypothesis test, we are “looking” for evidence against the null hypothesis. The null hypothesis can only be rejected if there is convincing evidence in the sample to support the alternative hypothesis. A hypothesis is always stated in terms of some population parameter (for now this parameter will be denoted μ). The exact formulation of the hypotheses depends on the aim of the analysis. The user will choose one of the following three formulations:

One-tailed hypotheses are stated in the following way

$$\begin{array}{ll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_A : \mu > \mu_0 & H_A : \mu < \mu_0 \end{array}$$

A two- tailed hypothesis takes the form

$$\begin{array}{l} H_0 : \mu = \mu_0 \\ H_A : \mu \neq \mu_0 \end{array}$$

In these formulations, μ_0 is a test value determined by the user.

Type I and type II errors

There is no guarantee that a true null hypothesis is not rejected. By the same logic, there is no guarantee that a false null hypothesis is rejected. Thus, we distinguish between two types of errors that may occur when performing a hypothesis test:

- *Type I error:* A type I error occur when the null hypothesis is rejected when it is true and should not have been rejected. As will be discussed below, the expected frequency of such errors can be controlled by the user (the individual carrying out the test).

- *Type II error*: A type II error occurs when a false null hypothesis is not rejected. Note that we do not have the same control over the type II errors.

Significance level α

It is desirable to minimize the occurrence of type I errors. Fortunately, we are able to control the expected frequency of such errors by choosing an appropriate significance level, here denoted α . When the chosen α is low, more statistical evidence against the null is required in order to reject it. On the other hand, when the chosen α is high, less evidence is needed. Typical values of α are 0.01, 0.05 or 0.1.

Below, we outline some frequently used tests.

One-sample t -test

The one-sample t -test can be employed to test whether or not the true population mean, denoted μ , is equal to, or different from, a specified test mean, denoted μ_0 . Depending on the research question, the test is either one-tailed or two-tailed. Formally, we state the one-tailed hypotheses in the following way

$$\begin{array}{ll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_A : \mu > \mu_0 & H_A : \mu < \mu_0 \end{array},$$

while the two-tailed hypotheses take the form

$$\begin{array}{l} H_0 : \mu = \mu_0 \\ H_A : \mu \neq \mu_0 \end{array}$$

The test statistic, also referred to as the t -statistic, is computed in the following way

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{SE(\bar{x})} \\ &= \frac{\bar{x} - \mu_0}{\frac{S_X}{\sqrt{n}}} \end{aligned}$$

where \bar{x} is the realized sample mean and μ_0 is the true mean under the null hypothesis. The expression in the denominator is the computed *standard error* associated with \bar{X} , *not to be confused with the sample standard deviation of X* . When the null hypothesis is true, and given certain assumptions (given below), it can be showed that the test statistic is t -distributed with $n - 1$ degrees of freedom.

One-sample t -test using critical values

Based on the chosen α , we need to find the *critical value* for the test. The critical value of a two-tailed test is $t_{\alpha/2}$, whereas the critical value of a one-tailed test is either t_α or $-t_\alpha$.

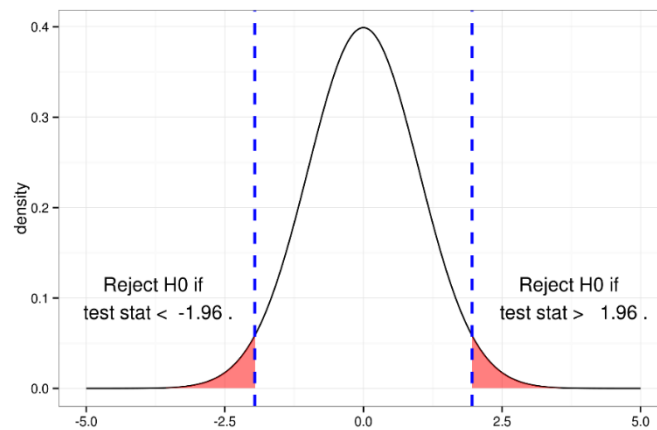
Rules for when to reject the null hypothesis are summarized in the following table

H_0	H_A	Reject H_0 when
$\mu = \mu_0$	$\mu > \mu_0$	$t > t_\alpha$
$\mu = \mu_0$	$\mu < \mu_0$	$t < -t_\alpha$
$\mu = \mu_0$	$\mu \neq \mu_0$	$ t > t_{\alpha/2}$

where t denotes the value of the test statistic that is computed from the realized sample.

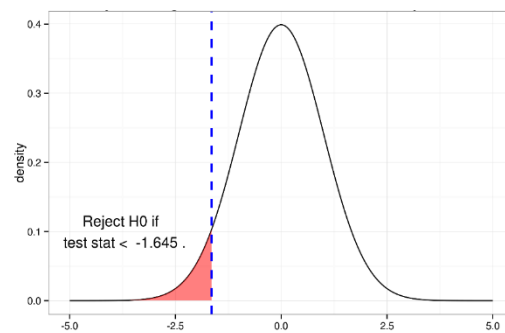
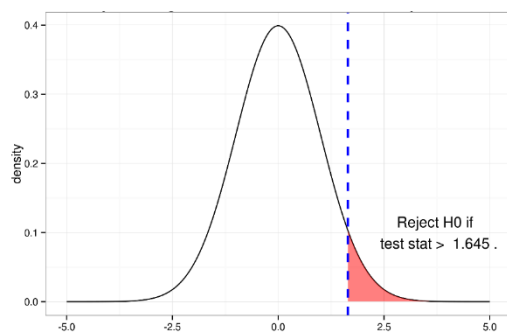
Graphically, the decision rules (or rejection regions) are depicted in the following way:

Regions of rejection for a two-tailed 5% hypothesis test (t -distribution $n > 100$)



For one-tailed tests, the 5% rejection region is located solely in one tail of the distribution.

Regions of rejection for a one-tailed 5% hypothesis test (t -distribution $n > 100$)



One-sample t -test using p -values

Instead of using critical values to determine when to reject the null-hypothesis, we may use the so-called p -value approach.

The p -value is defined as:

The p -value of a test is the probability of obtaining a test statistic as extreme, or more extreme, as the one realized from the sample, given a true null-hypothesis.

So, how to find the p -value? The following table summarize how to obtain the p -value for the three formulations of the alternative hypothesis:

H_0	H_A	p -value =
$\mu = \mu_0$	$\mu > \mu_0$	$P(T \geq t \mid H_0)$
$\mu = \mu_0$	$\mu < \mu_0$	$P(T \leq t \mid H_0)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$2 \cdot P(T \geq t \mid H_0)$

The rejection rule is

Reject the null hypothesis when the p -value $< \alpha$

One problem of using the p -value approach is that it generally requires computational power to obtain the p -value. In some special cases, we may be able to approximate the p -value using the t -table. Also, recall that when df tend to infinity (becomes a large number), the t -distribution tend to a standard normal distribution. Thus, we can use the normal table to find the p -value when n is large.

One-sample t -test using confidence intervals

We can easily relate confidence intervals to the one-sample t -test. Consider the following two-tailed hypothesis

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

The decision rule is that if μ_0 is not contained in the interval, the null-hypothesis is rejected. Note that our presentation is limited to only consider two-tailed hypothesis tests. One could expand the framework to also consider one-tailed hypothesis testing.

Assumptions:

For the one-sample t -test, the assumptions are:

1. The sample is random.
2. Distributional assumption:
 - a. *Small sample size*: if n is small, it must be assumed that the sample is drawn from a normal population (or at least approximately normal).
 - b. *Large sample size*: if n is large, the CLT applies and no distributional assumption is needed.

Example 2 (taken from <http://www.stat.columbia.edu/~martin/W2024/R2.pdf>)

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 Salmonella levels, as measured in MPN/g, were:

0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418

- a. Is there evidence that the mean level of Salmonella in the ice cream is greater than 0.3 MPN/g (for simplicity, let X denote MPN/g). Formulate a relevant hypothesis and carry out the test using $\alpha = 0.05$.
- b. What is the approximate p -value of the test in a.?

Two-sample t -test (independent samples)

Next, we consider the two-sample t -test. The aim of the analysis is to compare two populations means. Let μ_1 denote the mean of the first population and let μ_2 denote the mean of the second population. The research question dictates the formulation of the hypothesis.

Again, there are three different formulations to consider:

One-tailed:

$$\begin{array}{ll} H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 \\ H_A : \mu_1 > \mu_2 & H_A : \mu_1 < \mu_2 \end{array}$$

Two-tailed:

$$\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_A : \mu_1 \neq \mu_2 \end{array}$$

Let $X_{1,1}, X_{1,2}, \dots, X_{1,n_1}$ and $X_{2,1}, X_{2,2}, \dots, X_{2,n_2}$ be the two independent samples. From the data, we obtain the realized sample statistics $\bar{x}_1, \bar{x}_2, s_{\bar{x}_1}^2$ and $s_{\bar{x}_2}^2$. This in turn allows to compute the test statistic given by

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} \\ &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{\bar{x}_1}^2}{n_1} + \frac{s_{\bar{x}_2}^2}{n_2}}} \end{aligned}$$

The degrees of freedom df is obtained using the same expression as previously given when computing the confidence interval for the difference in means

$$df = \frac{\left(\frac{s_{\bar{x}_1}^2}{n_1} + \frac{s_{\bar{x}_2}^2}{n_2}\right)^2}{\frac{(s_{\bar{x}_1}^2/n_1)^2}{n_1 - 1} + \frac{(s_{\bar{x}_2}^2/n_2)^2}{n_2 - 1}}$$

(rounded down to the nearest integer value).

The decisions rules are summarized in the following table:

H_0	H_A	Reject H_0 when
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$t > t_\alpha$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$t < -t_\alpha$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$ t > t_{\alpha/2}$

As previously, t is the realized value of the test statistic. Alternatively, we can use the p -value approach. The rejection rule, and the way to obtain the p -value, is the same as stated above:

H_0	H_A	p -value =
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$P(T \geq t \mid H_0)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$P(T \leq t \mid H_0)$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$2 \cdot P(T \geq t \mid H_0)$

Assumptions:

For the independent samples t -test, we have following assumptions:

1. The samples are random.
2. The two groups, from which the two samples are taken, are independent.
3. Distributional assumption:
 - a. *Small sample size*: if n is small, it must be assumed that the two samples are drawn from normal populations (or at least approximately normal).
 - b. *Large sample size*: if n is large, the CLT applies and no distributional assumption is needed.

Example (taken from Triola)

People spend around 5 billion USD annually for the purchase of magnets used to treat a wide variety of pains. Researchers conducted a study to determine whether magnets are effective in treating back pain. Pain was measured using the so-called visual analog scale. The results from the study are given in the following summary

Reduction in pain level after magnet treatment (treatment group):

$$n_1 = 20, \bar{x}_1 = 0.49, s_{X_1} = 0.96$$

Reduction in pain level after placebo treatment (control group):

$$n_2 = 20, \bar{x}_2 = 0.44, s_{X_2} = 1.40$$

- a. Take an exploratory view point and perform a two-tailed hypothesis test for the difference in means. Use $\alpha = 0.10$.
- b. What is the approximate p -value of the test performed in a.?