# Hypothesis Testing : Other scenarios

We've looked so far at having to choose between one of two alternative hypothesis, where we know the probability distribution of observations under each of them.

In practice, we often face decision problems where this is not the case. We are often faced with a situation where the null hypothesis represents either some prior belief about the world, or some reasonable starting assumption about it. However, there is no sharply defined alternative hypothesis; instead, the alternative is simply the negation of the null.

Typically, we assume that the distribution of the observed data is known, under the null hypothesis. Sometimes it is known fully, and sometimes only partially, up to some unknown parameter.

The goal is to decide whether the observed data are consistent with the null hypothesis, or whether the null should be rejected.

## Examples

1. An inventor has proposed a new miracle cure for the common cold, & the regulatory body has to decide whether to license it.
The null hypothesis is that the proposed cure is no better than a placebo, i.e., that the mean number of days to recover from a cold is the same for patients administered either the drug or a placebo.
The regulatory body will only license the drug if the null hypothesis can be rejected.

2. An insurance company wants to know whether it incurs different risks on male & female drivers within a ~~spep~~ specified age range. It has data on the amount in claims paid out for each driver it has insured. The null hypothesis is that there is no difference.

3. Speed cameras were introduced on a stretch of road, and we want to determine if it has reduced the accident rate. The null hypothesis is that it has made no difference.

## General approach

The usual approach to the version of the hypothesis testing problem described in the last two pages goes as follows:

1. Identify a test statistic $T$ which can be computed from (i.e. is a ~~fuon~~ function of) the data, $T = T(x_1, x_2, \ldots, x_n)$
   - with slight abuse of notation, we use $T$ both for the function & its value
2. The choice of test statistic should be such that its probability distribution $F_0$, or density $f_0$, under the null hypothesis can be calculated explicitly.
3. Decide if the ~~test statistic~~ value obtained for the test statistic is sufficiently likely that the null hypothesis can be retained, or is so unlikely that it should be rejected. We shall shortly discuss how to make this judgement.

## Illustrative Examples

This approach to hypothesis testing has been well developed for a few specific scenarios which occur frequently in applications. We will study two of these scenarios.

1. Normal with known variance

$H_0$ : $X_1, X_2, \ldots, X_n$ are iid with a $\mathcal{N}(0, \sigma^2)$ distribution, $\sigma^2$ is known.

Note that we are assuming that the mean is unknown, but we hypothesize the value zero for it.

The variance is assumed to be known, and we will not be testing its value.

A natural candidate for a test statistic is

$$T' = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

A closely related & more convenient choice is

$$T = (X_1 + X_2 + \cdots + X_n) / \cancel{\left(\frac{\sigma}{\sqrt{n}}\right)} (\sigma\sqrt{n})$$

Note that $T$ has an $\mathcal{N}(0,1)$ distribution, under the null hypothesis.

Suppose we have observed data $x_1, x_2, \dots, x_n$ and computed $T = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} x_i$

How do we decide whether or not to reject the null hypothesis?

As $T$ has a $\mathcal{N}(0,1)$ distribution, values of $T$ close to $0$ are consistent with the null hypothesis, while values much bigger than $1$ or much smaller than $-1$ constitute strong evidence against it.

This suggests the following approach: fix a threshold $\tau$, and reject the null hypothesis $H_0$ if $|T| > \tau$.

The choice of $\tau$ will determine the false alarm probability:

$$\begin{aligned}
&\mathbb{P}(\text{reject } H_0 \mid H_0 = \text{true}) \\
&= \mathbb{P}(T > \tau) + \mathbb{P}(T < -\tau) \\
&= 2\,\mathbb{P}(T > \tau) = 2(1 - \Phi(\tau)),
\end{aligned}$$

where $\Phi$ is used to denote the cdf of the standard normal distribution (and $\phi$ is used to denote its density).

## Significance levels and p-values

It is customary to choose the threshold $\tau$ so that the false alarm probability is no more than a specified value, $\alpha$, which is called the significance level of the test.

Traditional choices of significance level are 10%, 5% and 1%, but your choice should be guided by the application more than by tradition!

So, how do we choose $\tau$ to obtain a test with a 5% significance level? We want:

$$\mathbb{P}(|T| > \tau \mid H_0 \text{ true}) = \alpha = 0.05, \text{ i.e.}$$

$$2(1 - \Phi(\tau)) = 0.05, \text{ or } \Phi(\tau) = 0.975$$

An alternative approach is to leave the decision of whether to reject the null hypothesis to the end user, and simply report the p-value, namely the probability of observing a value of the test statistic at least as extreme (very large or very small) as that observed. In other words,

$$\text{p-value}(T) = 2(1 - \Phi(T)),$$

in this example.

## Remarks

1. Note that there is only one type of error in this version of the hypothesis testing problem, namely that of rejecting $H_0$ when it is true. As there is no well-specified alternative $H_1$, it is not meaningful to speak of the probability of rejecting $H_1$ when it is true.

   Hence, we only seek to guarantee a bound on the false alarm probability. This bound is called the significance level.

2. This version of the problem is sometimes called <u>significance testing</u>, to distinguish it from the <u>hypothesis testing</u> problem of choosing between two well-specified alternatives. The application often guides which of the two versions is used.

The Wikipedia page on this topic has a good discussion of history and controversies in the development of this subject. These are by no means closed!

## 2. Normal with unknown variance

Suppose that $X_1, X_2, \cdots, X_n$ are iid Gaussian random variables, and that, under the null hypothesis, they have mean zero & unknown variance $\sigma^2$, i.e.,

$$H_0: \quad X_1, X_2, \ldots, X_n \text{ iid} \sim \mathcal{N}(0, \sigma^2),$$
$$\sigma^2 \text{ unknown.}$$

The alternative is that the mean is non-zero.

How should we test the null hypothesis?

We cannot use the same test statistic as before because $\sigma$ is unknown, and the choice $T = \frac{1}{n} \sum_{i=1}^{n} X_n$ would have a distribution that depended on the unknown parameter $\sigma^2$.

Is there some way we could use the sample variance, $\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$, where $\bar{X}$ is the sample mean, to get rid of this dependence?

It turns out the answer is yes.

# $t$ and $\chi^2$ distributions

Suppose $Z_1, Z_2, \ldots$ are iid standard normal random variables, i.e., with the $\mathcal{N}(0,1)$ distribution

Let $W = Z_1^2 + Z_2^2 + \ldots + Z_k^2$.

Then we say that $W$ has a $\chi^2$ distribution with $k$ degrees of freedom, denoted $\chi_k^2$.
In fact, the $\chi_k^2$ distribution is the same as the gamma distribution $\Gamma(\frac{k}{2}, \frac{1}{2})$, with shape parameter $k/2$ and scale parameter $1/2$.
This can be shown by computing their moment generating functions and showing that they are the same.

Let $\mathcal{U}$ and $V$ be independent random variables with $\mathcal{U} \sim \mathcal{N}(0,1)$ & $V \sim \chi_k^2$.
Let $W = \frac{\mathcal{U}}{\sqrt{V/k}}$. We say that $W$ has a t-distribution with $k$ degrees of freedom, denoted $W \sim t_k$

The t-distribution looks like the normal, but with heavier tails. It is symmetric around 0.
As $k \to \infty$, $t_k \to \mathcal{N}(0,1)$.

The relevance of the t distribution to our hypothesis testing problem is made clear by the following theorem.

---

<u>Theorem</u> :

Suppose $X_1, X_2, \ldots, X_n$ are iid $\sim \mathcal{N}(0, \sigma^2)$. Then,

1) $\mathcal{U} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i \sim \mathcal{N}(0,1)$

2) $V = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2_{n-1}$,

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean

3) $\mathcal{U}$ and $V$ are independent.

---

It follows from the theorem that $\mathcal{U}/\sqrt{V/(n-1)}$ has a $t_{n-1}$ distribution.

As we don't know $\sigma$, we can't calculate $\mathcal{U}$ and $V$. So how does this help?

Notice that we can calculate $\frac{\mathcal{U}}{\sqrt{V/(n-1)}}$ as the $\sigma$'s cancel!

Thus, the test statistic

$$T = \frac{\mathcal{U}}{\sqrt{V/(n-1)}} = \sqrt{\frac{n-1}{n}} \frac{\sum X_i}{\sqrt{\sum (X_i - \bar{X})^2}}$$

has a known distribution : $T \sim t_{n-1}$

## The $t$ test

We now apply this result to our hypothesis testing problem. Recall the null hypothesis

$$H_0 : X_1, X_2, \ldots \text{ are iid} \sim \mathcal{N}(0, \sigma^2),$$

where $\sigma^2$ is unknown.

If we observe data $x_1, x_2, \ldots, x_n$, we can compute $\overline{x} = \frac{1}{n}(x_1 + \ldots + x_n)$, & the test statistic

$$T = \sqrt{\frac{n-1}{n}} \frac{\Sigma x_i}{\sqrt{\Sigma (x_i - \overline{x})^2}}$$

As before, values of $T$ close to $0$ support the null hypothesis, and values far away are evidence against it. How far?

Given a significance level $\alpha \in (0,1)$, choose $\tau$ such that

$$\mathbb{P}(|T| > \tau) = \alpha, \quad \text{where } T \sim t_{n-1}$$

Such a $\tau$ can be found using tables of the cdf of $t$ distributions (very traditional) or a computer.

Alternatively, you can report a p-value,

$$p = \mathbb{P}(\underset{t_{n-1}}{\underset{\wr}{|T|}} > T(x_1, \ldots, x_n))$$

(Apologies for the horrible notation!)

## One-sided vs. two-sided tests

So far, we have implicitly been considering so-called two-sided tests, where the null $\mu = 0$ is being tested against the alternative $\mu \neq 0$.

Often, in applications, the relevant alternative will be of the form either $\mu > 0$ or $\mu < 0$. For instance, in the example of licensing a new drug, we want to test the null that it is no different from a placebo against the alternative that it is better. If it is worse than a placebo, it definitely shouldn't be licensed!

So we consider the problem of testing

$$H_0 : \mu = 0 \qquad \text{vs.} \qquad H_1 : \mu > 0$$

The approach is almost exactly the same as we have seen for the two-sided case, and we use the same test statistic.

The difference is only in the last step.

## Normal with known variance

Suppose we have calculated the test statistic

$$T_{obs} = \left(\Sigma_{i=1}^{n} x_i\right) \Big/ \sigma\sqrt{n} \qquad (\text{for } T\text{-observed})$$

which has a $\mathcal{N}(0,1)$ distribution under $H_0$.

Given a significance level $\alpha$, we reject $H_0$ if

$$\mathbb{P}(\mathcal{N}(0,1) > T_{obs}) \leq \alpha, \quad \text{i.e., if}$$

$$1 - \Phi(T_{obs}) \leq \alpha \quad \text{or} \quad \Phi(T_{obs} \geq 1-\alpha).$$

If the observed $T_{obs}$ is negative, it is evidence against $H_1$, so we don't reject $H_0$. Only large positive $T_{obs}$ can lead to rejecting $H_0$.

## Normal with unknown variance

Again we calculate the test statistic, which in this case is

$$T_{obs} = \sqrt{\frac{n-1}{n}} \; \frac{\Sigma x_i}{\sqrt{\Sigma (x_i - \bar{x})^2}} \qquad \text{where } \bar{x} = \frac{1}{n} \Sigma x_i .$$

To carry out the one-sided test at significance level $\alpha$, we reject $H_0$ if

$$\mathbb{P}(t_{n-1} > T_{obs}) \leq \alpha, \quad \text{i.e., if}$$

$$1 - F_{n-1}(T_{obs}) \leq \alpha \quad \text{or} \quad F_{n-1}(T_{obs}) \geq 1-\alpha,$$

where $F_{n-1}$ denotes the cdf of the $t$ distribution with $n-1$ degrees of freedom.

## Remarks

1. In applications, $H_0$ will usually take the form
$$H_0 : X_1, X_2, \dots \quad \text{iid } \mathcal{N}(\mu, \sigma^2),$$
where the value is specified, possibly non-zero. You should think of $\mu$ as representing normal behaviour (no pun intended) of your system or process.
The goal is to test whether some change has occurred, resulting in abnormal operation and requiring some intervention or further testing.

2. In many applications, we don't want to compare the mean of one population with a target, but compare the means of, say, two populations, to determine whether they are equal. We don't study this problem in this course.

3. The aim of the course is to give you some familiarity with commonly used methods, and ways of thinking, in statistics, rather than to provide a comprehensive knowledge base.