

# Chapter 9. Hypothesis testing

---

# 1. General definition and approach

---

## Hypothesis test ideas

$$N(\underline{\mu}, \sigma^2)$$

Motivating example: Suppose the annual returns of some asset over the last five years are 0.01, 0.03, -0.02, 0.03, -0.04. Does the asset has positive return on average (ignore the order and think the returns are i.i.d.)?

$$\bar{X} = 0.01$$

---

## Toy Example: Which mean?

Imagine we have a machine that produces random numbers but sometimes goes out of calibration. Assume that the machine in normal conditions produces random numbers that come from  $N(0, 1)$ . When the machine slips out of calibration, the random numbers come from  $N(1, 1)$ . Based on the value of one or more of these random numbers, how can we decide whether the machine is in calibration or not?

This question can be approached as a significance test. First, we might assume that the machine is in calibration (has mean 0), unless—  
we see otherwise based on the observed data.

In hypothesis testing, we have two hypotheses,  
in this case, the hypotheses are

$$H_0 : X \sim N(0, 1), \quad H_A : X \sim N(1, 1)$$

$H_0$  is called the **null hypothesis**, the  $H_A$  is the **alternative hypothesis**. Note the two are not exchangable. We also implicitly assume that one of the hypotheses is indeed true.

---

How do we decide whether to accept or reject  $H_0$ , the null hypothesis? The decision must be made based on the data.

---

For example: when we have one observation.  
accept  $H_0$  if  $X < 0.7$ , and reject if  $X \geq 0.7$

If we have multiple observations, we might decide to accept  $H_0$  if  $\bar{X} < 0.7$ .

Thus the decision to accept or reject is based on a summary of observed data,  $\bar{X}$ , which we call the **test statistics**.

---

Since we will reject  $H_0$  if  $\bar{X} \geq 0.7$ , the set  $[0.7, \infty)$  is called the **rejection region**.

$(-\infty, 0.7)$  accept  $H_0$

When a decision is made, mistakes can happen. If the null hypothesis is falsely rejected, it is a **type-I error**. If the null is false but it is falsely accepted, this is a **type-II error**. By abuse of term, we also refer to the probability that type-I/II error happens as the type-I/II error. E.g., the type-I error of the previous test is  $P(\underline{\bar{X}} > 0.7 | \bar{X} \sim N(0, 1)) = \underline{0.242}$ .

$$H_0 \Leftrightarrow \bar{X} \leq 0.7$$
$$\bar{X} < z$$

Back to machine example: The threshold 0.7 we chose is arbitrary. In principle we can choose any threshold  $t$ . Consider the case we only have one observation. The type-I error is

$$P(X > t | H_0) = \underbrace{1 - \Phi(t)}_{\text{with } t}$$

while the type-II error is

$$P(X < t | H_A) = \underbrace{\Phi(t-1)}_{\substack{X \sim N(1, 1) \\ = P(X < t-1) = \bar{\Phi}(t-1)}} \quad \text{with } t$$

$$P(X < t | H_A) = \underbrace{\Phi(t-1)}_{N(0, 1)} \quad \text{with } t$$

Thus the type-I error decreases with  $t$  while the type-II error increases with  $t$ , and we can make one arbitrarily small at the cost of the other type of error. This is a general phenomenon in hypothesis testing.

## Formal Structure

---

In statistics, a hypothesis is a statement about a population characteristic.

*population parameter*

Specifically, we make an assumption and then attempt to show that assumption leads to an absurdity or contradiction, hence the assumption is wrong.

## An Analogy

---

The Statistical Hypothesis Testing process can be compared very closely with a judicial trial.

Two Hypotheses are then created.

$H_0$ : Innocent      $H_A$ : Not Innocent (Guilt)

We take the defendant as innocent unless there is enough evidence against him.

## Common Hypotheses:

The form of the null hypothesis is

$$H_0: \text{population characteristic} = \text{hypothesized value}$$

$\mu = 0$

The alternative (or alternate) hypothesis will have one of the following three forms (composite hypothesis):

$$H_A: \text{population characteristic} > \text{hypothesized value}$$

$$H_A: \text{population characteristic} < \text{hypothesized value}$$

$$H_A: \text{population characteristic} \neq \text{hypothesized value}$$

Simple alternative hypothesis  $H_A: \text{population characteristic} = \text{another hypothesized value}$ , It can be regarded as one of the previous problem

$$H_0: \mu = 1$$

$$H_A: \mu > 0$$

$$H_A: \mu \neq 0$$

# Error

---

		Null Hypothesis
Decision	True	False
Accept $H_0$	No Error	Type II Error $\beta$
Reject $H_0$	Type I Error $\alpha$	No Error

## Procedure for performing test:

Given a 'test size/level'  $\alpha$ , choose rejection region to control type I error: for example, if you want to reject when  $X$  is big, then choose  $t$  such that

$$H_A: \mu > 0$$

$$P(X > t | H_0) = \alpha$$



$$H_0: \mu = 0$$

$$H_A: \mu > 0$$

$$P(X > t | H_0) = \alpha$$

$$P(X - \bar{X} > t - \bar{X} | H_0) \sim N(0,1)$$

$$H_A: \mu < 0$$

$$P(X < t | H_0) = \alpha$$

$$H_A: \mu \neq 0$$

$$P(|X| > t | H_0) = \alpha$$

# P-value

In this course, besides type-I and II errors, we also consider p-value:

$p\text{-value} =$

$P(\text{test statistics is the observed value or more extreme } | H_0)$

What “more extreme” means depends on the alternative hypothesis. It indicates observations more unlikely under the null, and more likely under the alternative.

The P-value is the probability, assuming that  $H_0$  is true, of obtaining a test statistic value at least as inconsistent with  $H_0$  as what actually resulted.

---

The significance level,  $\alpha$ , is actually the largest P-value tolerated for rejecting a true null hypothesis (how much evidence against  $H_0$  we require). This value is decided before conducting the test.

If the P-value ( $P \leq \alpha$ ), then we reject  $H_0$ .

If the P-value ( $P > \alpha$ ), then we fail to reject  $H_0$ .

P-val 0.08

$\alpha = 0.05$

$\alpha = 0.1$

For the random number machine example with one single observation. If the test statistic is  $X$  and the alternative is  $\mu = 1$  (or  $\mu > 0$ ). If the observed value is 0.8, what is the p-value? If the observed value is 2, what is the p-value? What if the alternative is  $\mu \neq 0$ ?

$$P(X > 0.8) = 0.212$$

$$P(X > 2) = 0.023$$

---

Two ways to make decision about the machine:

1. Given  $\alpha$ , such as 0.05, 0.01, reject  $H_0$  as soon as  $p-value \leq \alpha$
2. Given  $\alpha$ , Choose rejection region to control type I error:

$$P(X > t | H_0) = \alpha$$

## (Optional material)

Connection between p-value and type-I error:

Suppose we want the type-I error  $\alpha$  to be 0.05, we set the threshold  $t$  such that  $P(X > t|H_0) = 0.05$ , and the rejection region is  $[t, \infty)$ . Based on observed value  $x$ ,  $H_0$  is rejected if and only if  $x > t$ , i.e.  $P(X > x|H_0) < 0.05$ .

The advantage of using p-value is that you can save it and decide whether to accept or reject later.

---

# Slightly more general setup

---

$$H_0 : \theta = \theta_0, H_a : \theta \in \Omega_a$$

type I error (size):

$$\alpha = P(\text{rejecting } H_0 \text{ when } \theta = \theta_0)$$

type II error (is a function):

$$\beta(\theta_a) = P(\text{accepting } H_0 \text{ when } \theta = \theta_a \in \Omega_a)$$

Power (Power Function):

$$power(\theta_a) = P(\text{rejecting } H_0 \text{ when } \theta = \theta_a \in \Omega_a) =$$

$$1 - \beta(\theta_a)$$



$$\mu = 0, \sigma^2$$

---

Example: Consider the machine example. Suppose we reject  $H_0$  if  $\bar{X} > c$ . What value of  $c$  should be used for size  $\alpha$  test?  ~~$\bar{X} \sim N(\mu, \sigma^2)$~~

$$P(\bar{X} > c \mid X_i \sim N(0, 1)) = \alpha$$
$$\bar{X} \sim N\left(0, \frac{1}{n}\right) \Rightarrow \sqrt{n}\bar{X} \sim N(0, 1)$$
$$P(\underbrace{\sqrt{n}\bar{X}}_{N(0, 1)} > \sqrt{n}c) = \alpha \Rightarrow \sqrt{n}c = z_\alpha$$
$$c = \frac{z_\alpha}{\sqrt{n}}$$



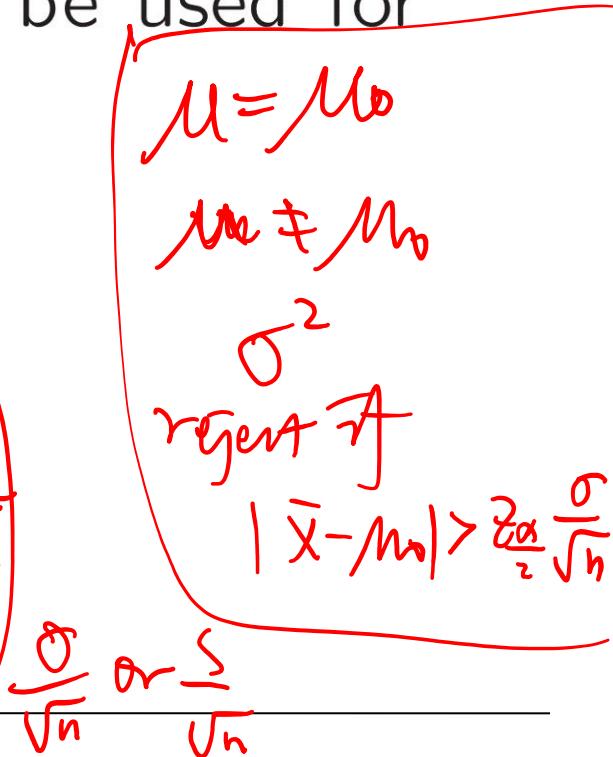
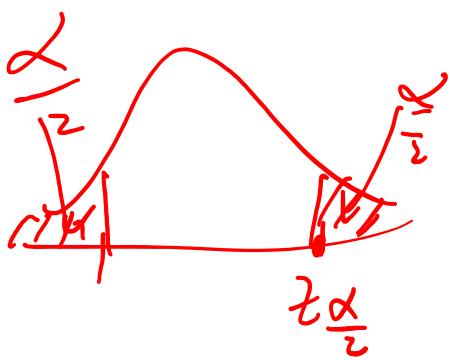
**Example:** Consider the machine example. But now we have the alternative hypothesis  $H_a$  :  $\mu \neq 0$ . Now it is more natural to reject  $H_0$  if  $|\bar{X}| > c$ . What value of  $c$  should be used for size  $\alpha$  test?

$$P(|\bar{X}| > c | X_i \sim N(0, 1)) = \alpha$$

$$= P(|N(0, 1)| > \sqrt{n}c | H_0) = \alpha$$

$$\Rightarrow \sqrt{n}c = z_{\frac{\alpha}{2}}$$

reject if  $|\bar{X}| > z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$



Let  $X_1, \dots, X_n \sim N(\mu, 1)$ . Want to test  $H_0 : \mu = 0$  vs.  $H_a : \mu > 0$ . Consider the test: reject  $H_0$  if  $\bar{X} > c$ .

$$c = z_{\alpha} \frac{1}{\sqrt{n}}$$

What value of  $c$  makes the test have size  $\alpha$ ?

What is the power function?

$$\begin{aligned} \text{power}(\mu) &= P\left(\bar{X} > z_{\alpha} \frac{1}{\sqrt{n}} \mid X_i \sim N(\mu, 1)\right), \quad \bar{X} \sim N\left(\mu, \frac{1}{n}\right) \\ &= P\left(\underbrace{\sqrt{n}(\bar{X} - \mu)}_{N(0, 1)} > \sqrt{n}\left(z_{\alpha} \frac{1}{\sqrt{n}} - \mu\right)\right) \\ &= P\left(N(0, 1) > z_{\alpha} - \sqrt{n}\mu\right) = 1 - \Phi(z_{\alpha} - \sqrt{n}\mu) \\ &= \Phi(\sqrt{n}\mu - z_{\alpha}) \end{aligned}$$

## 2. test for mean

---

### test of the mean:

The hypotheses take the form

$$H_0 : \mu = \mu_0, H_A : \mu < \mu_0, \mu > \mu_0, \text{ or } \mu \neq \mu_0$$

A natural test statistics is

$$T = \frac{\bar{X} - E(\bar{X}|H_0)}{SE(\bar{X}|H_0)} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Note the expected value and the standard error are found under the null.

In the case of normally distributed data, the sampling distribution of  $T$  under the null is known to be  $t(n - 1)$ . If  $n$  is large enough,  $T$  is approximately standard normal.

---

---

## Decision based on Type I error:

Suppose the alternative test is two sided:  $H_A : \mu \neq \mu_0$ . We reject  $H_0$  in favor of  $H_A$  if  $|T|$  is big. That is we reject  $H_0$  if  $|T| > threshold$ , and we set threshold to be  $z_{\alpha/2}$  or  $t_{\alpha/2}$  so type I error is  $\alpha$ .

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, \quad \text{if } \mu < \mu_0: \\ p\text{-value} = P(\bar{X} < \bar{x}) = P(T < t)$$

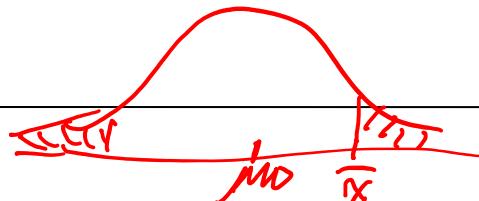

---

Let  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$  be the observed value of the test statistics, the p-value is computed by

$$p\text{-value} = \begin{cases} P(T \leq t | H_0) & \text{if } H_A: \mu < \mu_0 \\ P(T \geq t | H_0) & \text{if } H_A: \mu > \mu_0 \\ P(|T| \geq |t| | H_0) & \text{if } H_A: \mu \neq \mu_0 \end{cases}$$

$$\mu < \mu_0$$

$$P(|\bar{X} - \mu_0| > |\bar{x} - \mu_0|)$$



---

**Example:** A consumer group wishes to see whether the actual mileage of a new SUV matches the advertised 17 miles per gallon. The group suspects it is lower. To test the claim, the group fills the SUV's tank and records the mileage. This is repeated ten times. The results are

(11.4, 13.1, 14.7, 14.7, 15.0, 15.5, 15.6, 15.9, 16.0, 16.8)

Setting up hypotheses:

$$H_0 : \mu = 17$$

$$H_A : \mu < 17$$

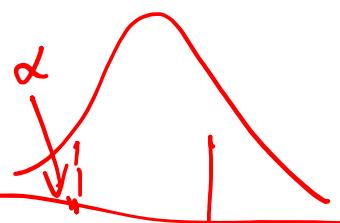
$\mu < \mu_0$

reject if  $\bar{X} < ?$

$$P(\bar{X} < ?) = \alpha$$

reject if  $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < ?$

$$P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -z_\alpha\right) = \alpha$$

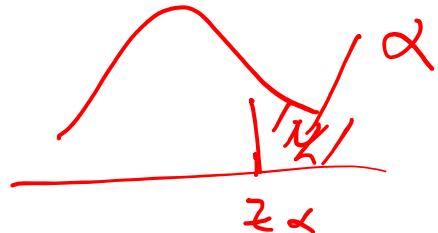


reject if  $\bar{X} < \mu_0 - z_\alpha \frac{S}{\sqrt{n}}$

$\mu > \mu_0$

reject if  $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > ?$

$$P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > z_\alpha\right) = \alpha$$

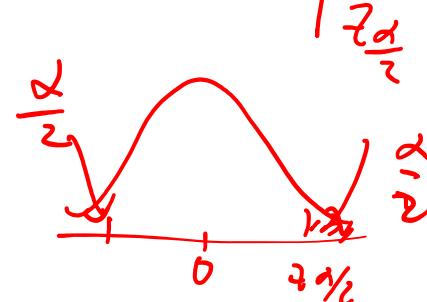


reject if  $\bar{X} > \mu_0 + z_\alpha \frac{S}{\sqrt{n}}$

$\mu \neq \mu_0$

reject if  $|\bar{X} - \mu_0| > ?$

$$P\left(|\frac{\bar{X} - \mu_0}{S/\sqrt{n}}| > ?\right) = \alpha$$



reject if

$$|\bar{X} - \mu_0| > \frac{z_\alpha}{2} \frac{S}{\sqrt{n}}$$

$\Leftrightarrow \bar{X} > \mu_0 + \frac{z_\alpha}{2} \frac{S}{\sqrt{n}}$  or

$$\bar{X} < \mu_0 - \frac{z_\alpha}{2} \frac{S}{\sqrt{n}}$$

$$\bar{x} = 14.87, S = 1.572, n = 10$$

$$p\text{-val} = P(\bar{X} < \bar{x}) = P\left(N(0, 1) < \frac{14.87 - 17}{1.572/\sqrt{10}}\right) \xrightarrow{\substack{N(0, 1) \\ \bar{T}_9}} 9.14 \times 10^{-6}$$

reject if  $\bar{x} < 17 - t_{\alpha} \frac{1.572}{\sqrt{10}} \xrightarrow{\substack{N(0, 1) \\ \bar{T}_9}} 16.185$   
 $t_{q, \alpha} \xrightarrow{\substack{\\ \bar{T}_9}} 16.088 \Rightarrow \text{reject } H_0$   
if  $\alpha = 0.05$