Motivation: In 1000 tosses of a coin, 560 heads and 440 tails appear. The maximum likelihood estimation gives

\[ \hat{\theta} = 0.56. \]

Is it reasonable to assume that the coin is fair?
A hypothesis is a statement about a population parameter.

The two complementary hypothesis in a hypothesis testing problem are
- the null hypothesis, denoted by $H_0$
- the alternative hypothesis, denoted by $H_1$

Example
- The population parameter is $\theta$
- Null hypothesis: it is a fair coin $H_0 : \theta = \frac{1}{2}$
- Alternative hypothesis: it is not a fair coin $H_1 : \theta \neq \frac{1}{2}$
Example continued

Suppose $\theta_0 = \frac{1}{2}$ (if null hypothesis is true), then $\overline{X}_n$ satisfies

$$
\frac{\overline{X}_n - \theta_0}{\sqrt{\theta_0(1 - \theta_0)/n}} \xrightarrow{d} \mathcal{N}(0, 1).
$$

In other words, with probability $1 - \alpha$, $\overline{X}_n$ should fall into

$$
\left( \theta_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1 - \theta_0)}{n}}, \theta_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1 - \theta_0)}{n}} \right)
$$

If $\theta_0 = \frac{1}{2}$, $n = 1000$, $\alpha = 0.05$, and $z_{0.975} = 1.96$,

$$
(0.4690, 0.5310)
$$

In other words, 0.56 does not fall into this interval (it is not confidence interval)! We should reject the null hypothesis $H_0 : \theta_0 = \frac{1}{2}$!
How to decide which hypothesis we should accept/reject?

We reject $H_0$ if the observed value of $T(X)$ belongs to a region $R \subseteq \mathcal{X}$.

The statistics $T(X)$ is called the test statistics and $R$ is called the rejection region or critical region.
Test statistics and rejection region

In the coin tossing example: $T(X) = \overline{X}_n$ and the rejection region is

$$R = \left( -\infty, \theta_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n}} \right) \cup \left( \theta_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}, \infty \right)$$

where $\theta_0 = \frac{1}{2}$ is the parameter in $\Theta_0$.

Equivalently, we reject $H_0 : \theta = \theta_0$ if

$$\left| \frac{\overline{X}_n - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \right| > z_{1-\frac{\alpha}{2}}.$$

where $\overline{X}_n$ is the observed value of $\overline{X}_n$.

However, is it possible that $T(X)$ and the choice of $R$ give you a wrong answer?
Type I and II error

There are two types of errors:

- **Type I error**: we reject $H_0$ but $H_0$ is the truth.
- **Type II error**: we retain $H_0$ but $H_1$ is the truth.

**Table:** Summary of outcomes of hypothesis testing

<table>
<thead>
<tr>
<th></th>
<th>Retain Null $H_0$</th>
<th>Reject Null $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ true</td>
<td>✓</td>
<td>Type I error</td>
</tr>
<tr>
<td>$H_1$ true</td>
<td>Type II error</td>
<td>✓</td>
</tr>
</tbody>
</table>

It is important to note that reject $H_0$ does not mean accept $H_1$. 
Power function

The power function of a test with rejection region $R$ is defined by

$$
\beta(\theta) = \mathbb{P}_\theta(T(X) \in R)
$$

where $X \sim F(x; \theta)$.

**Remark:** The power function is the probability of rejecting $\theta$. 
Size of a test

The size of a test is defined to be

\[ \alpha = \sup_{\theta \in \Theta_0} \beta(\theta) \]

where \( \Theta_0 \) consists of all parameters in the null hypothesis. A test is said to have level \( \alpha \) if its size is less than or equal to \( \alpha \).

- The size of a test is the maximal probability of rejecting the null hypothesis when the null hypothesis is true.
- If the level \( \alpha \) is small, it means type I error is small.
Example

Let $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ where $\sigma$ is known. We want to test

$$H_0 : \mu < 0 \text{ versus } H_1 : \mu > 0$$

Hence

$$\Theta_0 = (-\infty, 0] \text{ versus } \Theta_1 = (0, \infty).$$

Note that

$$T(X) = \bar{X}_n$$

is the MLE of $\mu$. We reject $H_0$ if $T(X) > c$ where $c$ is a number.
We reject $H_0$ if $T(X) > c$. The power function is

$$
\beta(\mu) = \mathbb{P}_\mu(T(X) > c) = \mathbb{P}_\mu(\bar{X}_n > c) \\
= \mathbb{P}_\mu \left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} > \frac{\sqrt{n}(c - \mu)}{\sigma} \right) \\
= 1 - \Phi \left( \frac{\sqrt{n}(c - \mu)}{\sigma} \right)
$$

What is the size? Note that $\beta(\mu)$ is increasing!

$$
\sup_{\mu \leq 0} \beta(\mu) = \beta(0) = 1 - \Phi \left( \frac{\sqrt{nc}}{\sigma} \right).
$$

To have a size $\alpha$ test, we set $\beta(0) = \alpha$:

$$
c = \frac{\sigma \Phi^{-1}(1 - \alpha)}{\sqrt{n}} = \frac{\sigma z_{1-\alpha}}{\sqrt{n}}
$$
Here $\beta(0) = \alpha = 0.05$, i.e., the probability of rejecting the null hypothesis when $H_0$ is true is smaller than 0.05.
The Wald test

Consider testing

\[ H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \]

Assume that \( \hat{\theta} \) is asymptotically normal:

\[ \frac{\hat{\theta} - \theta_0}{\hat{se}} \sim \mathcal{N}(0, 1) \]

if \( \theta_0 \) is the true parameter. Then the size \( \alpha \) Wald test is: reject \( H_0 \) when

\[ |W| \geq z_{1 - \frac{\alpha}{2}} \]

where

\[ W = \frac{\hat{\theta} - \theta_0}{\hat{se}} \]

is close to standard normal and \( \hat{se} \) is the estimated standard deviation of \( \hat{\theta} \).
For example, in the coin tossing example,

\[ \hat{\theta} = \bar{X}_n \]

and

\[ \hat{se} = \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}. \]

We reject \( H_0 : \theta_0 = \frac{1}{2} \) if

\[ |\hat{\theta} - \theta_0| \geq z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}. \]