# Introduction to Statistical inference

# Procedure of statistical analysis

1. Set the goal of statistical analysis
2. Collect data
3. Make a model
4. Fit the model using collected data
5. Make statistical inference

# Some terminologies

- Unit: A single unit, usually an object or person, whose characteristics are of interest
- Population: The complete collection of units about which information is sought
- Random variable: A numerical summary of the characteristics of a unit in the population
- (parametric) model for random variable $X$:
  $\mathcal{F} = \{f(x; \theta) : \theta \in \Omega\}$
- Parameter space $\Omega$: a set of plausible values of parameter $\theta$
- Random sample: a set of independent and identically distributed random variables, $X_1, \ldots, X_n$

# Goal of statistical inference

- Statistical model contains unknown parameter which represents the information we want to know.
- Unless we collect all pupulation data, we never know the true parameter $\theta$.
- We only have a finte random sample from the population.
- Statisical inference includes how we can draw information for the unknown true $\theta$ from the random sample

# Estimation

- How to find or estimate the true $\theta$ from a given random sample?

- Statistic: Suppose that $n$ random variables $X_1, \ldots, X_n$ constitute a sample from the distribution of a random variables $X$. Then any function $T = T(X_1, \ldots, X_n)$ of the sample is called a Statitic.

- Point estimator: A statistic to make a guess for $\theta$

- Interval estimator: a plausible interval for the true parameter $\theta$

- What is the best way to estimate $\theta$? In what sense?

- We need to consider Bias and Variance of our estimator

- The most ideal estimator for $\theta$ is the estimator that is unbiased and has the smallest variance among all unbiased estiamtor.

- Sometimes we need to make decision between two choices based on a given random sample
- Similar to the estimation, we should handle the uncertainty of a random sample
- Our question is how to make the most scientifically ideal decision and how to explain the uncertainty of your decision

## Motivating example

- Suppose that we have a coin and we want to test whether or not this coin is fair. To get sample, we flip a coin six times and get results.

- Suppose we have three Heads and three Tails. Can we say a given coin is fair?

- Suppose we have six Heads. Can we say a given coin is fair?

## Terminologies

- Hypothesis: A statement about parameter
- Null hypothesis ($H_0$): A hypothesis that the parameter takes a particular value.
- Alternative hypothesis ($H_1$): A hypothesis that the parameter falls in some alternative range of values.
- Type I error: $H_0$ is rejected when $H_0$ is true
- Type II error: $H_0$ is not rejected when $H_0$ is true
- Type I error probability ($\alpha$): Probability that Type I error occurs
- Type II error probability ($\beta$): Probability that Type II error occurs

# Type I error and Type II error

- Type I error probability ($\alpha$): The probability that we reject $H_0$ when $H_0$ is true.

- Type II error probability ($\beta$): The probability that we accept $H_0$ when $H_0$ is not true.

|          |       | True          |               |
|----------|-------|---------------|---------------|
|          |       | $H_0$         | $H_1$         |
|          | $H_0$ | No error      | Type II error |
| Decision |       |               |               |
|          | $H_1$ | Type I error  | No error      |

Now our question is

- What is the best test?
- How to construct the best test?