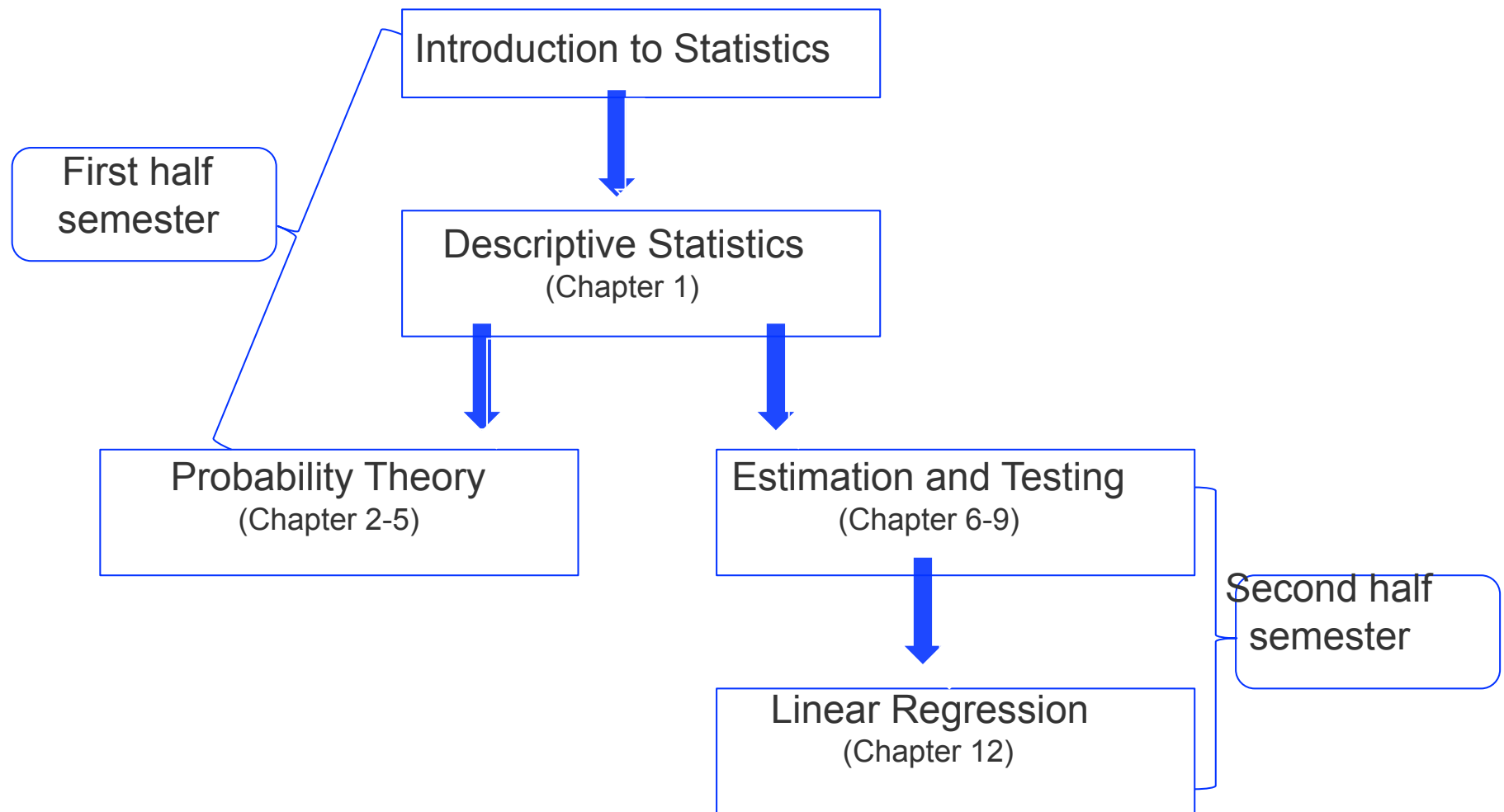


# Overview of the course



# Basic concepts

- **Population:** the whole class of individuals which an investigator is interested in.
- **Census:** the desired information is available for all objects in the population.
- **Sample:** a subset (part) of the population which is examined or observed.
- **Sample Size:** the number of observations in a single sample.
- **Variable:** any characteristic whose value may change from one object to another in the population, including *univariate*, *bivariate*, *multivariate*.

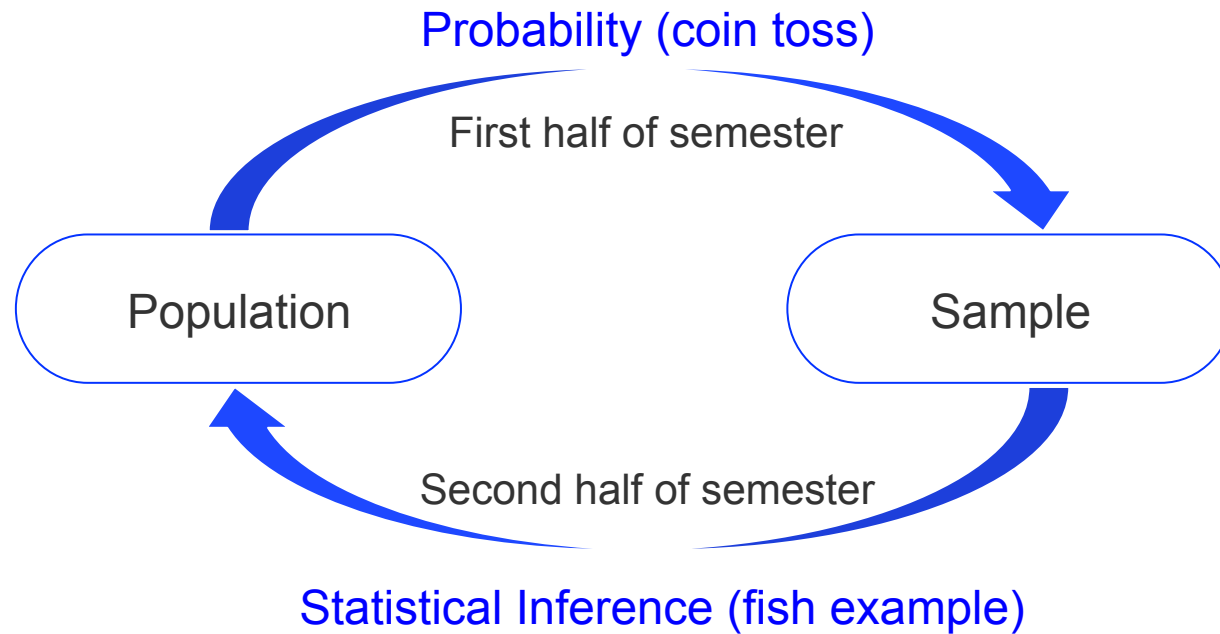
# Probability

- What are random variables? Example: coin tosses.
- To describe random variables: *distribution*. This course will cover a variety of commonly used probability distributions.
  - Discrete distributions: Binomial, Poisson, etc.
  - Continuous distributions: Exponential, Normal (Gaussian), etc.
- Conditional probability.

# Statistical Inference

- Estimation:
  - Point estimation. Example: What is the total number of fish in a lake?
  - Interval estimation.
- Hypothesis testing:
  - One sample testing.
  - Two sample testing. Example: Is there a significant improvement in the new drug?
- Estimation and hypothesis testing are just two different ways of looking at the same problem.

# Probability and Inference



# Examining a new data set

1. Examine each variable by itself.
2. Study the relationship between variables.

For both steps 1 and 2 we want to:

- Display the data graphically.
- Summarize the data numerically (Statistics).
- Construct a mathematical model.

# Descriptive Statistics

- Pictorial methods:
  - Stem-and-Leaf Displays.
  - Dotplots.
  - Histograms.
- All these methods convey information about the following aspects of the data:
  - Identification of a typical or representative value
  - Extent of spread about the typical value
  - Presence of any gaps in the data
  - Extent of symmetry in the distribution of values
  - Number and location of peaks
  - Presence of any outlying values

# Stem-and-Leaf displays

- Steps for constructing a Stem-and-Leaf Display:
  1. Select one or more leading digits for the stem values. The trailing digits become the leaves.
  2. List possible stem values in a vertical column.
  3. Record the leaf for every observation beside the corresponding stem value.
  4. **Indicate the units for stems and leaves someplace in the display.**
- R demo for Stem-and-Leaf:
  - Command: `>stem(x)`
  - Option: `scale=...`, `scale` has to be a positive number. It controls the plot length. A value of `scale=2` will cause the plot to be roughly twice as long as the default (`=1`).



# More basic concepts

- **Discrete Variable:** Its set of possible values is either finite or else can be listed in an infinite sequence. (Gender, Age, etc.)
- **Continuous Variable:** Its possible values consist of an entire interval on the real number line. (Height, Weight, etc.)
- **Frequency:** Number of times a value occurs in the data set.
- **Relative Frequency:**  $\text{Frequency} / (\text{Sample size})$ .