

# Dis 1: General Info; Data; Population vs. Sample

Related textbook chapters: 1 and 4

Ch 1 and 4 handout and solution offered by Dr. Pac can be accessed here: [Handout](#) [Solution](#)

This handout incorporates reviews with all exercises from the handout given by Dr. Pac.

## 1 General Info

### 1. Contact Me

You can reach me by **sending me an email** or **attending my office hours**.

- Email me at [travis.cao@wisc.edu](mailto:travis.cao@wisc.edu) (please start the subject line with “Econ 310”).
- Office hours take place at the following times and locations:
  - Mondays, 9:30 - 10:30am, online via Zoom (link on Canvas)
  - Thursdays, 1:00 - 2:00pm, in person @ 7226 Social Sciences
  - Or by appointment

### 2. Discussion Sections

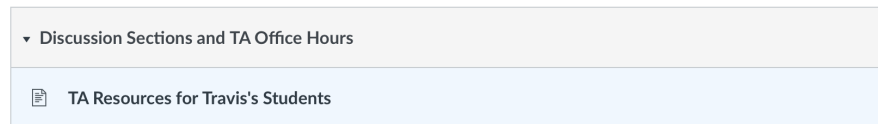
- Attendance
  - Attendance is not required, but strongly encouraged.
  - Sections take place at the following times and locations:
    - \* Thursdays, 2:25 - 3:15pm @ 586 Van Hise
    - \* Thursdays, 3:30 - 4:20pm @ 590 Van Hise
    - \* Thursdays, 4:35 - 5:25pm @ 590 Van Hise

Due to the classroom capacity limit, please attend the section that you registered for.

- Screen recording of the discussion section will be posted (I use my iPad to teach all sections, and my screen will be recorded and posted on Canvas after my last section of the week; all sections cover the same material, so only one section recording will be posted each week).
  - \* **I encourage you to attend discussion section in person every week**, since screen recording is, at best, an imperfect substitute for attending sections live.
  - \* Screen recording is intended as a resource to allow you to re-watch part of the section in case you didn’t follow along at the time, or because you cannot make it to certain week’s section due to any personal / health reason.
  - \* Bottom line: You all are adults. Make smart choices.

- Discussion handout
  - If you notice the box underneath the title of this handout, you’ll see that this handout differs from the version given by Dr. Pac.
    - \* I personally like to spend some time in sections to review the concepts that we learned in the past week, and then spend the remaining time going through exercises.
    - \* Dr. Pac’s version of handout under the section “Lab Session Review Worksheets” on Canvas is great, since all the exercises come from him directly. But I often like to review the relevant material in a slightly different way to help you more easily digest the content, so I created this version of handout that you’re currently holding.

- \* My handout contains the same exercises and reviews the same set of concepts as Dr. Pac's version, but adds more material to help us efficiently and effectively go through the concepts together.
- Bottom line: My version of handout is the only set of handout that you need for this course!
- Accessing discussion materials online
  - Handouts, solutions, section iPad screen recording, and any discussion material (including online office hour's Zoom link) are available on Canvas.
  - You can find them on "TA Resources for Travis's Students", which is under the module "Discussion Sections and TA Office Hours" (located at the bottom of your Canvas 310 course home page).



## 2 Data

- What are we studying in this class about "statistics"?
  - "Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of **data**." – Wikipedia
  - Data is at the core of the study of statistics, so let's first take a high level view of
    - \* How to categorize different types of data, and
    - \* What exactly can one do with a set of data
- **How to categorize different types of data?**
  - Commonly, data can be categorized based on the values recorded, or based on how much data points are collected.
  - If we categorize data **based on the values recorded**, we can divide data into 3 types:
    1. **Interval data**: The values recorded are actual numbers that make meaningful sense.  
e.g.
    2. **Ordinal data**: The values recorded represent a ranked order.  
e.g.
    3. **Nominal data**: The values recorded are arbitrary (typically only used as identifier).  
e.g.
  - If we categorize data **based on how much data points are collected**, we can divide data into 2 types:
    1. **Population**: A set of data that records all items of interest.
    2. **Sample**: A set of data that records only a subset of items of interest.

e.g. For everyone sitting in the front row, their favorite numbers are:

- **What exactly can one do with a set of data?**

1. **Descriptive statistics:** A set of methods used to summarize or present your data.  
e.g. Making a bar graph from your data  
e.g. Calculating the mean (average) of your data
2. **Inferential statistics:** A set of methods used to draw conclusion or make inference about the population using a sample data.  
e.g. Say a sample from all Econ 310 students has been collected, and within the sample, 82% of them are sophomores.  
Now, when asked to estimate the percentage of all students in Econ 310 that are sophomores, you might guess \_\_\_\_\_.

### 3 Population vs. Sample

- We just mentioned that population and sample data differs based on how much data points are collected, and that two sets of methods – descriptive and inferential statistics – can be used to describe your data.
- Obviously, if one always has access to population data, then inferential statistics seem rather meaningless: you already have the population data, so there's no need to make inference about the population from a sample.
  - But, as you can intuitively see, this is likely not going to be the case: population data often is much harder to get, which is why inferential statistics matter.
  - Inferential statistics will be a big part of what we study for the rest of this semester. It's a harder set of methods compared with descriptive statistics (think about how can one tell that the inference made about the population makes sense), but it's more useful.
- For this first week, instead of looking at the more complex inferential statistics, let's look at some descriptive statistics that you can use.

<b>Descriptive statistics for a population (Parameter)</b>	<b>Descriptive statistics for a sample (Statistic)</b>
Population median	Sample median
Population mode	Sample mode
Population mean $\mu = \frac{1}{N} \sum_{i=1}^N x_i$	Sample mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

- Question: Why do we sometimes refer to the mean of  $x$  as  $\mu$ , and sometimes as  $\bar{x}$ ?

## 4 Exercises

1. A manufacturer claims that 1% of the artificial hearts it has ever produced are defective. When 1,000 hearts are randomly drawn, 1.5% are found to be defective.
  - (a) What is the population of interest?
  - (b) What is the sample?
  - (c) What is the parameter?
  - (d) What is the statistic?
2. You are shown a coin. The owner of the coin claims it's "fair" (meaning that it will produce the same number of tails and heads when flipped a very large number of times).
  - (a) Describe an experiment to test the claim?

For the rest of the question, suppose we do this 100 times.

(b) What is the population in your experiment?

(c) What is the sample?

(d) What is the parameter?

(e) What is the statistic?

(f) Recall your goal is to determine whether the coin is fair. What conclusion would you draw if 99 of the 100 flips came up heads? What conclusion would you draw if 50 of the 100 flips came up heads?

3. Consider grade data for the following sample of students (drawn randomly from the entire population of 350 students who took Econ 310 last semester):

Student	Grade
Tom	80
Sean	90
Ed	60
Ben	70
Nate	80

(a) What are  $N$  and  $n$ ? Describe the difference between the two.

(b) What are  $\bar{x}$  and  $\mu$ ? Describe the difference between the two.

(c) Calculate the median, mode, and range.

4. Ten people in a room have an average height of 5 feet 6 inches. An 11th person, who is 6 feet 5 inches tall, enters the room. Find the average height of all 11 people?

Note: 1 foot = 12 inches