M358K: September 9th, 2020.

What is statistics?

1st Identify the broad population of interest about which you're trying to draw conclusions:

e.g.,
- cats;
- voters in the US;
- math majors;
- snakes;
- bacteria in Petri dish;
- Harry Potter fans;
- Policy holders;
- Coronavirus infected
- middle schools in AISD

2nd What are you trying to measure?

e.g.; color, breed, hrs of sleep, ...

- age, how they vote, education, ...
- length, venomous or not, ...
- how many survived after a week of treatments, ...
- house, ...
- number of claims, ...
- age, where, recovery time, ...
- # of students w/ free lunches, ...

3rd The members of your population which you measure or survey:

CASES or **OBSERVATIONAL UNITS**

↕

LABEL

4th For every case we measure a set of quantities or outcomes:

VARIABLE

- **Quantitative (Numerical)**
  - **Discrete**
    - # of children
    - # of cars owned
  - **Continuous**
    - height
    - weight
- **Categorical (Qualitative)**
  - **Nominal**
    - major
    - favorite color
  - **Ordinal**
    - letter grades
    - Starfleet Ranks

Focus: Figuring out a model for relative likelihoods of a particular variable.

Named Distributions w/ **parameters**

e.g., Binomial (n, p)

usually the parameter of interest;

Poisson (λ)

Pareto

Normal (mean $= \mu$, var $= \sigma^2$)

*** Even w/out a name for the dist'n, we can be interested in the average behavior, i.e., the estimate of the mean. ***

5 $\stackrel{+4}{=}$ Collect your data and analyze your data: to figure out an appropriate SHAPE of the dist'n in the model and/or values of interest (estimates).