Goals of Statistics

Populations and Samples

Population
- Set of subjects, items, or events we want to learn about
- Generally very large or infinite
- All people, all men/women, all pigeons, all concrete or abstract nouns

Sample
- Subset of population assessed in a given study
- Much smaller
- Randomly selected
- Not perfectly representative of population (sampling variability)

Sometimes population is hypothetical
- Experimental trials (repetitions)
- Your reaction time to a red square
  - Imagine we could repeat infinite times
  - Sample is finite times we actually do it

Random Selection
- Every member of population must have equal chance of inclusion
- Property of data-gathering process
  - Study design must take random selection into account
- Otherwise sample is biased
  - Only testing students at library
- Selection variables may interact with outcome
  - People/events in sample may differ from rest of population
    - Undergrads
    - Racial distribution

Parameter
- Characteristic of the population
- Usually theoretically meaningful
- Mean, variance, proportion, rate, correlation
  - What’s the average IQ of college students?
  - How many attempts does it take a normal rat to learn this maze?
  - How many attempts if we cut out its hippocampus?
  - What fraction of words can a subject remember?
  - What’s the correlation between height and extraversion?

Statistic
- Mathematical function to be computed from data
- Difference between statistic and its value
  - E.g., mean is a statistic (arithmetic average)
  - Value for any dataset will be some number
- Usually serves one of three functions
  - Descriptive statistic: Summarizes some aspect of the sample data
  - Estimator: Estimates some parameter of the population
  - Inferential statistic: Aids testing of some hypothesis about the population
Descriptive Statistic
Summarizes some aspect of the data
  Mean, median, maximum, quartiles, standard deviation, etc.
Used only for describing sample data
  Not for making inferences about population
  Can be first step of data analysis
  Also useful if sample is all you’re interested in
E.g. average age of students in class

Estimator
Estimates some parameter of the population
  Mathematical function applied to sample that usually gives close to correct answer for population
Usually also a descriptive statistic
  Sample mean is estimator for population mean
  Difference is just in what you use it for

Sampling error
  Value of estimator almost never exactly correct
  Different samples give different results

Inferential Statistic
Aids testing of some hypothesis about the population
Indicates how reliable an effect in the sample is
Value generally has no physical meaning
  Not like inches, time, or even psychological variables
Examples: $t$, $F$, $\chi^2$, $p$