Overview

- Part 1 - Introduction: What is Statistics?
- Part 2 - Basic Concepts: Variables & measurement scales
- Part 3 - Graphical displays & shapes of distributions

Part 1: Introduction

What is “statistics”?  

- Statistics refers to statistical methods & procedures...
- and to the results/outputs of those methods
- Averaging is a statistical method and the result of averaging data in a sample — i.e., the mean — is a statistic.
Role of statistics in research

Statistics helps us to collect, organize, summarize, analyze, interpret, & present data

“Statistics is the grammar of science.” - Karl Pearson

“...the purpose of statistics is to organize a useful argument from quantitative evidence, using a form of principled rhetoric*.” - Robert P. Abelson

*rhetoric: the art of effective/persuasive speaking or writing

Role of statistics in research

“The role of statistics is not to discover truth. The role of statistics is to resolve disagreements between people.” - Milton Friedman

Role of statistics in research (PPDAC)
Understanding & defining the problem of interest
Statistics is used in many contexts

- Statistics & data-based decision making go hand-in-hand
- Statistics plays vital roles in:
  - many areas of research in engineering, natural sciences, social sciences, & humanities
  - machine learning & data mining; artificial intelligence
  - development of public policy in a wide range of areas

Statistics is the science of us, and that is why we should be fascinated by numbers.
Many Ways of Collecting Data

- Two ways we will consider in this course:
  - Correlational Studies -
    - measure associations between predictor & criterion variables
  - “subjects” come with their own set of variables
- Designed Experiments -
  - measure effects of independent variables on dependent variables
  - random assignment of “subjects” to experimental conditions

Data Management & “Cleaning”
Statistics plays important roles at beginning of studies

- Brain-imaging studies often use only right-handed adults participants
  - what is an “adult”? how should we measure “handedness”?
- UK study compared survival rates of children who had heart surgery
  - what is a “child”? what counts as “heart surgery”? when can a death be attributed to the surgery?
- Political surveys often measure preferences of likely voters... how should we define or measure a “likely” voter?

Descriptive Statistics
Data Summary & Communication

- describes/summarizes important characteristics of data
- uses graphs & statistics e.g., mean or standard deviation
- describes interesting features of sample

Populations vs. Samples
(inferential vs. descriptive statistics)

- Population: all events (subjects, scores, etc) of interest
- Sample: subset of population
  - random sample: each member of population has equal chance of being selected
  - convenience sample (e.g., psychology undergraduates)
- Can we infer characteristics of population from sample?
Inferential Statistics

- uses sample to make claims about a population
  - e.g., estimate population parameters from sample statistics
  - e.g., investigate differences among population by examining differences among groups/samples
- Psychology experiments typically test undergraduate students & generalize results to all young adults

Descriptive & Inferential Statistics

- Population parameters $\mu$, $\sigma^2$ (Inferential Statistics)
- Statistical theory e.g., sampling distributions
- Sample statistics mean, variance $\bar{Y}$, $s^2$ (Descriptive Statistics)

Exploratory vs Confirmatory Analyses

- Exploratory Data Analysis
  - first major proponent was John Tukey
  - goal: discover & summarize interesting aspects of data
  - discover interesting hypotheses to test
- Confirmatory Data Analysis
  - data are gathered & analyzed to evaluate specific a priori hypotheses
  - example: clinical drug trials
- Important not to confuse two types of analyses
  - replication crisis in Psychology related to confusion about two types of research

Part 1 Summary

- what is statistics?
- why study statistics?
- ways of collecting data
  - designed experiments vs. correlational studies
  - causal effects vs. associations
- descriptive vs inferential statistics
  - pop vs sample
  - params vs statistics
- exploratory vs confirmatory statistics