Chapter 1: Stats Starts Here

Section 1.1: What is Statistics?

Statistics: the science of data, including how to collect, classify, summarize, organize, analyze, and interpret data and how to draw conclusions or make generalizations from data. (learning from data)
(field of study / discipline)

Statistics is used in:

- Agriculture

  - how much omega 3 to feed a chicken, so that the eggs contain as much omega 3 as possible, but still taste good.

- Business

  - insurance rates

  - advertising : who to target, best way to reach them
    ⤷ online info / activity

- Games/Sports

  - betting odds

  - teams: • roster decisions (who plays together, against other teams
            • trades

- Health/Medicine

  - drug trials : effectiveness / safety of a drug

            • compare drugs

- Politics

  - polls : predicit election results, gauge public opinion

  - call election results (winners)

  - government : predict income / expenditures policy decisions resource allocation.

Three main aspects of Statistics:

- **Design:** how to obtain data to answer questions. ch 9,10
  - Surveys
  - experiments } randomization
  - observational studies
- **Description:** methods for summarizing/displaying the data. Descriptive
  - Summary values : averages   Statistics
  - graphs : bar charts, histograms

- **Inference:** making decisions and predictions based on the data.
  - Regression   Inferential
  - ANOVA test   Statistics

Section 1.2: Data

Data: a collection of numbers, characters, images, or other items that provide information about something.

Data requires context:

- **Who:** cases or individuals for which (or whom) we have recorded data. person / thing you want information about

- **What:** characteristics/attributes being measured/recorded.
  → Variables

- **Where:** where the data were recorded.

- **When:** when the data were recorded.

- **Why:** why the data were recorded. What do you want to know/accomplish?

- **How:** how the data were collected/measured. → randomization
  done incorrectly → incorrect conclusions

Example: A potato chip company produces salt and vinegar chips at one of its factories. Every night at this factor, they maintain quality control by randomly selecting 20 bags of chips from that day's production. The bags are weighed and the results are recorded in grams.

- Who: 20 bags of salt and vinegar chips

- What: weight (in grams)

- Where: a factory

- When: every night

- Why: quality control: checking that the weight of each bag is "close" to weight printed on bag.

- How: random sample.

Terminology

Data values are often referred to as **observations**.

Data Table: a table used to organize data.

- rows correspond to the individual cases about which (or whom) we have collected data. "who"

- columns correspond to the characteristics that have been recorded in each case. "what", variables