Math 201

Statistics

The Science of Statistics

Statistics is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, presenting, and interpreting numerical and categorical information.

Are people good intuitive statisticians?

An individual has been described by a neighbor as follows: "Steve is very shy and withdrawn, invariably helpful but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structurut and stre, and a passion for detail." Is Steve more likely to be a librarian or a farmer?

- The resemblance of Steve's personality to that of a stereotypical librarian strikes everyone immediately.
- But, there are more than 20 male farmers for each male librarian!
- Bayes' theorem

Are people good intuitive statisticians?

Consider the letter N. Is N more likely to appear as the first letter in a word or as the third letter?

- Scrabble players: First letter
- In fact: N occurs more frequently in the third position

Are people good intuitive statisticians?

A certain town has two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

(i) The larger hospital

(ii) The smaller hospital

(iii) About the same (that is, within 5% of each other)

- Intuitive answer: (iii)
- Actually: (ii)
- Law of Large Numbers

Are people good intuitive statisticians?

For further discussion and additional examples:

Thinking, Fast and Slow by Daniel Kahneman

Why We Learn/Study Statistics

- Analyze news and data critically
- Avoid manipulation and misinformation
- Statistical thinking is a core part of critical thinking
- Probably very useful to your future career
- Maybe useful in your hobby (eg. sports)

Why We Learn/Study Statistics

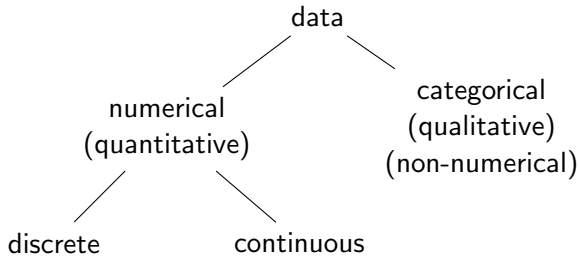
Statistics plays a central role in many fields, including:

- Business & Economics: finance, marketing, operations
- Social Sciences: psychology, sociology, political science
- Health & Medicine: clinical trials, epidemiology, public health
- Engineering & Technology: quality control, reliability
- Natural Sciences: biology, environmental studies, physics
- Computer Science: machine learning

Chapter 1

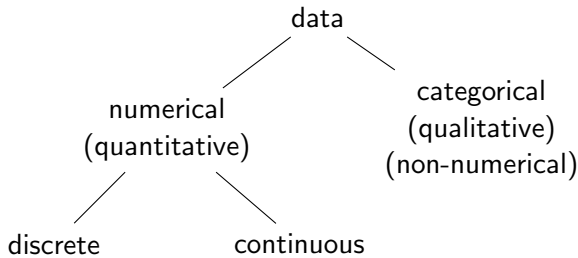
Introduction to Statistics

Types of Data



- Numerical (Quantitative) Data consists of numerical values representing counts or measurements.
  - Discrete Numerical Data: countable
  - Continuous Numerical Data: measurable
- Categorical (Qualitative) Data consists of categorical values describing qualities or attributes.

Types of Data



- Discrete numerical data examples: shoe size, family size, letter grade, ...
- Continuous numerical data examples: temperature, age, height, weight, ...
- Non-numerical data examples: hair color, employment status (employed, unemployed, retired), political affiliation ...

Collecting Data

- Generally, you can obtain data in three different ways:
  - From a published source
    - book, journal, newspaper, web, etc.
  - From a designed experiment
    - the researcher exerts strict control over the units (people, objects, or things) in the study
    - treatment group vs control group
    - example: drug studies
  - From an observational study
    - researcher observes the experimental units in their natural setting and records the variable(s) of interest
    - example: surveys
    - example: a study about smoking

Types of Statistical Applications

- Descriptive statistics utilizes numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present that information in a convenient form.
  - methods for summarizing and describing a data set you already have
  - describe the sample/data you saw
- Inferential statistics utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.
  - methods for using a sample to draw conclusions about a larger population, with uncertainty explicitly accounted for
  - infer something about the population you didn't fully observe

Descriptive statistics

- What it does: turns raw data into a clear picture: "What does the data look like?"
- Typical outputs:
  - Numbers: mean, median, mode, min/max, range, variance, standard deviation, percentiles
  - Tables/plots: histograms, bar charts, boxplots, scatterplots
- Example: In a class, the average exam score is 62, the median is 65, and scores range from 30 to 98. And you might show a histogram or boxplot to visualize this data.

Inferential statistics

- What it does: answers "what can we conclude and how sure are we?"
- Typical tools:
  - Estimation: confidence intervals (e.g., "population mean is likely between 68 and 74")
  - Hypothesis testing: p-values, significance tests (z-test, t-test, chi-square test, ANOVA, etc.)
- Example: From a random sample of 50 voters, we estimate the candidate's support is $52\% \pm 6\%$.

Sampling

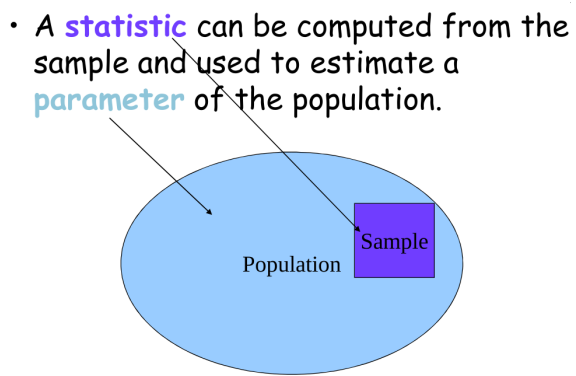
- An experimental (or observational) unit is an object (e.g., person, thing, transaction, or event) about which we collect data.
  - example: one student
- A variable is a characteristic or property of an individual experimental (or observational) unit in the population.
  - example: exam score

Sampling

- A population is a set of all units (usually people, objects, transactions, or events) that we are interested in studying.
- A sample is a subset of the units of a population.

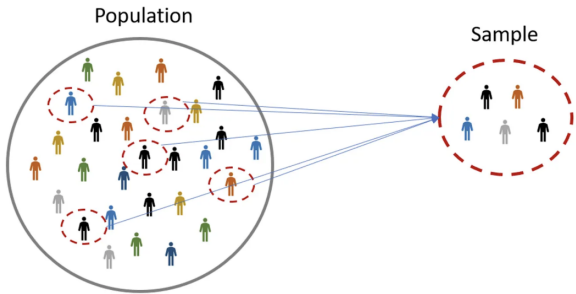
Sampling

- A statistic is a numerical summary calculated from a sample.
  - what investigators know
  - example: sample average
- A parameter is a numerical summary describing a population.
  - It is usually unknown
  - what investigators want to know
  - example: population average



Sampling

- A sample should accurately represent the population; therefore, it must be selected using an appropriate random sampling method.



- Random sampling is a key idea that allows us to obtain representative samples from a population.

Random Sampling

- The theory and practice of random sampling are rich enough to fill an entire semester-long course in statistics or survey methodology.
- In MATH 201, we will not go into the details of random sampling. Instead, we only mention a simple but fundamental method:
  - A simple random sample of n experimental units is a sample selected from the population in such a way that every different sample of size n has an equal chance of selection.
- In addition to simple random samples, there are more complex random sampling designs such as:
  - stratified random sampling, cluster sampling, systematic sampling, randomized response sampling, etc.

Example: Starbucks Price Perception

- Starbucks Türkiye management wants to understand how customers perceive its pricing strategy.
- Since sales receipts reveal what customers buy but not what they think about prices, the company conducts a short survey.
- Using an appropriate sampling methodology, a total of 300 customers across 20 randomly selected stores are asked:
  - Do you think Starbucks is reasonably priced?
- The responses (Yes/No) are recorded to estimate the proportion of customers who believe Starbucks offers reasonable prices.

Example: Starbucks Price Perception

- Experimental Unit: An individual customer in the survey
- Variable of Interest: Response to "Do you think Starbucks is reasonably priced?" (Yes/No)
- Population: All Starbucks customers in Türkiye
- Sample: The 300 customers surveyed in 20 randomly selected stores
- Parameter: The true proportion of all Starbucks Türkiye customers who think Starbucks is reasonably priced
- Statistic: The proportion of the 300 surveyed customers who answered "Yes" (e.g., 62%)

Example: Starbucks Price Perception

- The purpose of the Starbucks price perception study is to understand customer perceptions of pricing. Specifically:
  - Estimate the proportion of customers who believe Starbucks is reasonably priced
  - Provide insights that sales data cannot reveal — how customers feel about pricing
  - Inform management decisions on pricing strategy, loyalty, and competitive positioning
- Pricing is one of the most critical business decisions a company can make. It affects customer perception, demand, and profitability.

Statistics and Computation

- Statistics deals with challenging real-world problems and relies heavily on computation.
- Modern statistical analysis requires strong software support.
- Common tools include: R, Python, SPSS, Excel.
- In MATH 201, we will use Excel.
  - Excel is not designed specifically for statistics, but it provides strong tools for descriptive analysis and basic inference.
  - It is widely available, and most students already have basic familiarity. This makes Excel an excellent choice for an introductory statistics course.

Coming Up in MATH 201

- Part 1 - Descriptive Statistics
  - Numerical measures (mean, median, standard deviation, ...)
  - Graphical methods (histograms, box plots, ...)
- Part 2 - Probability
  - Random variables
  - Normal distribution
  - Central limit theorem
- Part 3 - Inferential Statistics
  - Confidence intervals
  - Hypothesis tests