

1. INTRO TO DATA

DESCRIPTIVE STATS

INFERENTIAL STATS



eZcZ Stat Network



eZcZ Stat Network

Statistics Made EZ

1. DATA

Variation

- *Refers to differences in a characteristic among individuals or items; variation can also refer to fluctuation over time.*



eZcZ Stat Network
Statistics Made EZ

Examples

- Day-to-day stock variation
- NFL 40-yard dash times.
- Housing prices in a given city.

A bit on variation

- Variation is everywhere (different magnitudes).
- Both qualitative and quantitative variables reveal variability in data.



Variability is what makes statistics so interesting and allows us to interpret, mode, and make predictions from data.

Individuals/Observational Units

- Subjects/objects of the population of interest; can be people, sports teams, machines, or anything that we want to study.

Variable

- Any characteristic of an individual that we are interested in. Qualitative and quantitative.

Variable

- Observed value of a variable on a specific individual.

Individuals/Observational Units

- Subjects/objects of the population of interest; can be people, sports teams, machines, or anything that we want to study.

Variable

- Any characteristic of an individual that we are interested in. Qualitative and quantitative.

Variable

- Observed value of a variable on a specific individual.

- Variables are denoted by a capital letter (i.e., X).
- Use capital letters to indicated that we have NOT seen/observed their values.
- The value of X are still undetermined and the outcome of the observation depends on some *random process*
→ Random variables
- Once we see an observation can use lowercase.



EXAMPLE

- X = the grade on your next statistics exam.
- x = your observed grade on the exam.

KNOW THE DIFFERENCE between a random variable and an observation (i.e., data point).



Time Series Data



- Data collected over time.
- Common for financial data.
- Plot against time.
- Known as a time series.

Categorical Variables

- Individuals can be placed into distinct categories, nominal and ordinal variables.
- Nominal: order not meaningful: colors, weather, etc.
- Ordinal: grades, educational degrees, etc.



Quantitative Variables

- Take numerical values for which arithmetic operations make sense.



eZcZ Stat Network
Statistics Made EZ

Examples

- Height/weight
- Amount of precipitation
- Etc.

Descriptive Statistics

- Collection, presentation, and description of data in form of graphs, tables, and numerical summaries that provide meaningful information about the data.



eZcZ Stat Network
Statistics Made EZ

Look for pattern and summarize/present data.

Types of Descriptive Statistics

Better understand a variable's

- Distribution
- Variability
- Central tendency



Inferential Statistics

Deals with the interpretation of data as well as drawing conclusions and making generalizations based on data for a larger group of subjects (i.e., a population).



eZcZ Stat Network
Statistics Made EZ

- Make data-based decisions.
- Generalize information obtained from descriptive analysis to a larger group of individuals.

Descriptive vs. Inferential Statistics

Example: Prior to the 2019 NFL season a randomly selected group of fans from across the US were asked to provide an overall rating for their feeling on the new pass interference rule resulting in the following:



**30% very satisfied, 20% satisfied, 15% indifferent
17% dissatisfied, and 18% very dissatisfied.**

Descriptive vs. Inferential Statistics

30% of the fans interviewed were very satisfied with the new pass interference rule. → Descriptive



We can expect about 30% of all NFL fans to be very satisfied with the new pass interference rule. → Inferential

Population

The entire group of individuals or objects about which we want to gain information.



eZcZ Stat Network
Statistics Made EZ

* All currently enrolled students from University X in a study that investigates average student debt of students at University X.

Population

Defined by the target group of interest. Changes when the target group and purpose of the study change.



eZcZ Stat Network
Statistics Made EZ

- Who is it that we want to learn about?

Process

Component of a system that has inputs and outputs (vague).



eZcZ Stat Network
Statistics Made EZ

- Annual income for statistics professors in New Jersey.
- Monthly sales for a startup company in California.

Sample

Part of a population, usually randomly selected, from which we obtain information in order to draw conclusions about the entire population/process.



- Every 5th Apple customer at a particular Apple store.
- Random sample of students from University X.

Statistics vs. Parameters

- Any numerical summary describing a sample is called a statistic.
- Any numerical summary describing a population is called a population parameter.



Example: Salary of all NFL players

- Mean salary of all NFL players: population parameter.
- Mean salary of a random sample of 50 NFL players: sample statistic.

Statistics vs. Parameters

- Any numerical summary describing a sample is called a statistic.
- Any numerical summary describing a population is called a population parameter.

Example: Salary of all NFL players

- Mean salary of all NFL players: population parameter.
- Mean salary of a random sample of 50 NFL players: sample statistic.



Extra Definitions

Distribution → collection of all possible values the variable can take and how often each value occurs.



Frequency table → summary that shows the distribution of a variable.

Categorical data: bar charts, pie charts, pareto charts, etc.

Quantitative data: histogram.



eZcZ Stat Network
Statistics Made EZ

UP NEXT....
DESCRIBING CATEGORICAL
DATA