

Summary of Week 1

Goals of statistics are to **understand the world** and **describe or make predictions** about it.

Vocabulary:

Descriptive Statistics – describe world
make predictions

- **Population** – collection of all items of interest
- **Parameter** – number describing population in some way

Inferential Statistics – understand world

- **Sample** – subcollection of population which will be studied (hopefully “*representative*” of population)
- **Statistic** – number computed from sample, hopefully approximating a parameter

(Random) variable is something you can *measure* or *observe*
– written with capital letter **X** or **Y**

The value you measure or observe is called **data** (or “a datum”) -- written with lower-case **x** or **y**

Making a measurement or observation is called an **experiment** (or **trial**)

Categorical Variable \longleftrightarrow **Qualitative Data**

(Non-numerical description / classification)

- Can compute most / least frequent, proportions
- Make bar graphs, pie charts

Ordinal data

has natural ordering
(e.g. “low” “medium” “high”)

Nominal data

no natural ordering
(e.g. “red” “yellow” “blue”)

Numerical Variable \longleftrightarrow **Quantitative Data**

(Numbers)

- Can compute max / min, average, quartiles, variance
- Make scatterplots, regression lines, histograms

Discrete data

from counting
(gaps between possible values)

Continuous data

from measuring
(nearby values count as “same”)

Example. You want to investigate the average hair length of UVI students.

You walk into the library and start measuring hair length of everyone you see.

- **Population:** all students at UVI
- **Parameter:** average hair length on UVI
- **Sample:** students in the library
- **Statistic:** average hair length in library
- **Variable:** length of a student's hair in inches
- **Data:** (numbers like 3, 15, etc)

* If the sample of students in the library isn't representative of the UVI population then we may not get a good estimate.

For example, men and women may have different average hair length. The library may have a higher proportion of men than the university.

Example. You want to investigate whether cats who eat more also nap more.

You get your neighbors to make measurements of their cats' eating and napping habits.

- **Population:** all cats in the world!
- **Parameter:** correlation between eating and napping
- **Sample:** cats of neighbors
- **Statistic:** correlation between eating and napping of neighbor cats
- **Variables:** X=amt of food, Y=amt of nap time, Cor(X,Y)
- **Data:** (numbers like 3, 15, etc)

* If the sample isn't **representative** then we may not get a good estimate.

* If the sample isn't **big enough** we may not get a good estimate.

* If values have **error**, we may not get a good estimate.