Chapter 1

Professor Tim Busken

January 14, 2014
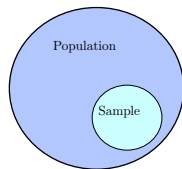
Table of Contents

Descriptive and Inferential Statistics

Definition

Descriptive Statistics consists of procedures used to summarize and describe the important characteristics of a set of measurements.
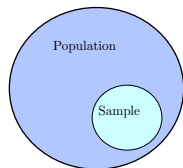
Descriptive and Inferential Statistics

Definition
Descriptive Statistics consists of procedures used to summarize and describe the important characteristics of a set of measurements.
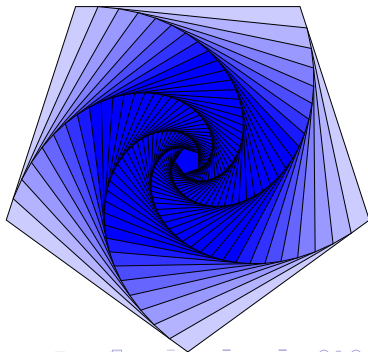
Definition
Inferential Statistics consists of procedures used to derive logical conclusions about population characteristics from sample data.

Data

Definition
Data are collections of observations (such as measurements, genders, survey responses).

Statistics

Definition
Statistics is the science of planning studies and experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data
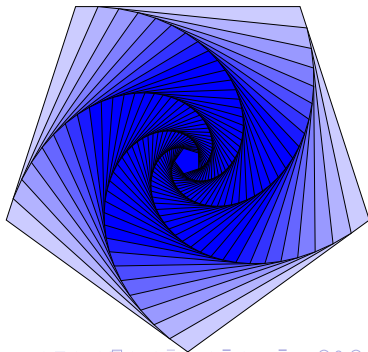
Population

Definition

Population: the complete collection of all individuals (scores, people, measurements, and so on) to be studied; the collection is complete in the sense that it includes **all** of the individuals to be studied.
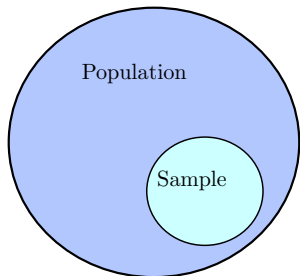
Sample

Definition

A Sample is a subcollection of members selected from a population
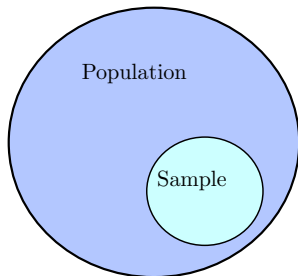
Census

Definition
A Census is a collection of data from every member
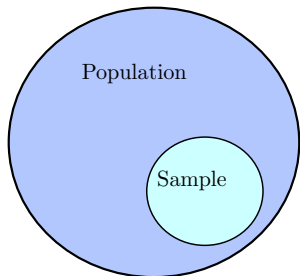of a population
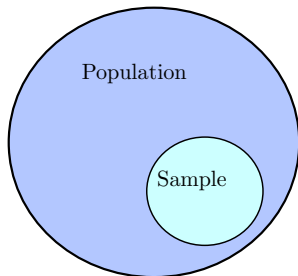
Census

Definition

A Census is a collection of data from every member of a population

It is often difficult or impossible to take a census (or gather census data) because the population is too big and resources are limited.

Sample Data

Definition

Sample Data is a collection of data from a subset of a population

Sample Data

Definition
Sample Data is a collection of data from a subset of
a population

A goal of statisticians is to draw reasonable
conclusions about population characteristics based
off of the analysis of sample data.   It is extremely
important to obtain sample data that are
representative of the population from which the data
are drawn. [1]

Guidelines

- Sample data must be collected in an appropriate way, such as through a process of random selection.

- If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical torturing can salvage them.
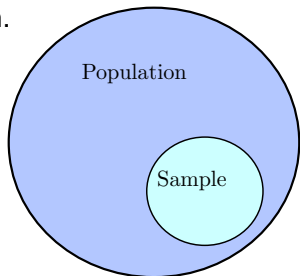
Statistical Thinking

- <u>Context of the data</u>: Is it relevant?

- <u>Source of the data</u>: Where did you get the data?

- <u>Sampling method</u>: How did you get obtain the data?

- <u>Conclusions</u>: Does sample data tell you anything about the population?

- Practical implications: Does the inference based off of sample data suggest action be taken to change policy?
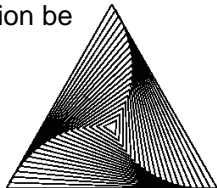
Statistical Significance

- Consider the likelihood of getting the results by chance.

- If results could easily occur by chance, then they are not statistically significant.

- If the likelihood of getting the results is so small, then the results are statistically significant.
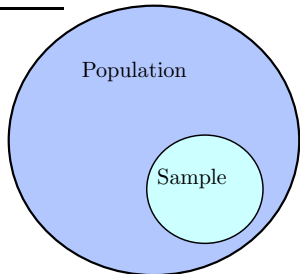
Parameter or Statistic?

Definition
A parameter is a numerical measurement describing some characteristic of a population.

For instance, the average height of <u>all</u> women is a (population) parameter.

Parameter or Statistic?

Definition
A statistic is a a numerical measurement describing some characteristic of a sample.

For instance, the average height of <u>all</u> women <u>in this class</u> is (sample) statistic.
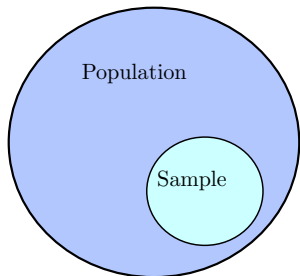
Variables and Experimental Units

Definition

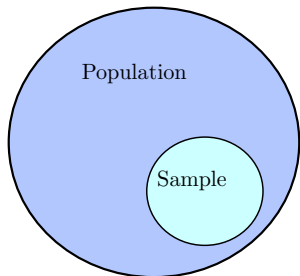A variable is a characteristic that changes or varies over time and/or for different individuals or objects under consideration.

Definition

An experimental unit is the individual or object on which a variable is measured. A single measurement or data value results when a variable is actually measured on an experimental unit.

Example

A sample of four students from our college was taken, and the following measurements were recorded in the table below.

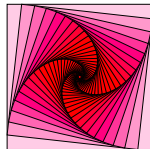| Student	GPA	Gender	Year	Major	Current Number of Units Enrolled
1	3.2	F	Fr	Accounting	16
2	2.2	M	So	Mathematics	12
3	3.4	M	Fr	Nursing	12
4	2.0	F	Fr	Business	15

Observations

- The students are the experimental units of the study.

- The measurements in the table are associated with the five variables: GPA, gender, year, major and current number of units enrolled.
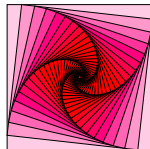
Definition
Univariate data result when a single variable is measured on a single experimental unit.

Definition
Bivariate data result when a two variables are measured on a single experimental unit.

Definition
Multivariate data result when more than two variables are measured.

Example

A sample of four students from our college was taken, and the following measurements were recorded in the table below.

| Student	GPA	Gender	Year	Major	Current Number of Units Enrolled
1	3.2	F	Fr	Accounting	16
2	2.2	M	So	Mathematics	12
3	3.4	M	Fr	Nursing	12
4	2.0	F	Fr	Business	15

Observations

- The table contains multivariate data.

Is the Data Qualitative or Quantitative?

Definition

Quantitative data consists of numbers representing counts or measurements.

For instance, the weights of supermodels or the ages of respondents.
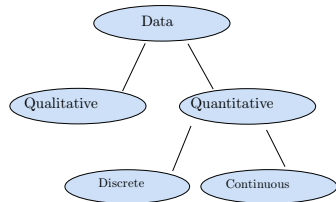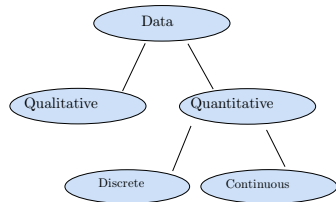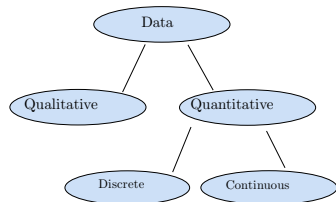
Is the Data Qualitative or Quantitative?

Definition

Qualitative data consists of names or labels
(representing categories).

For instance, the colors in the color wheel.

Quantitative data can further be described by
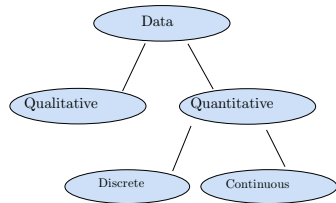distinguishing between discrete and continuous types.

Discrete data

Definition
Discrete data result when the number of possible
values is either a finite number or a countable number
(i.e. the number of possible values is 0, 1, 2, 3, . . .)

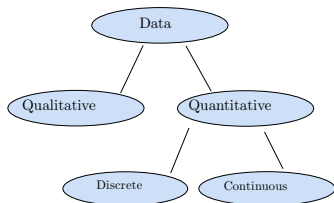Example: The number of eggs that a hen lays.

Continuous data

Definition
Continuous data result from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps

Example: The amount of milk that a cow produces; e.g. 2.343115 gallons per day

Levels of Measurement [1]

Ratio	There is a natural zero starting point and ratios are meaningful	Example: Distances
Interval	Differences are meaningful, but there is no natural zero starting point and ratios are meaningless.	Example: Body temps.
Ordinal	Categories are ordered, but differences can't be found or are meaningless.	Example: Letter grades, A, B, C, D, F.
Nominal	Categories only. Data cannot be arranged in an ordering scheme.	Example: eye colors.
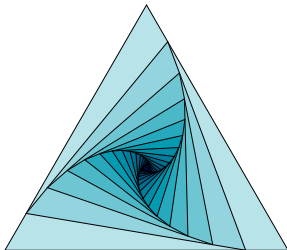
M. F. TRIOLA, *Essentials of Statistics*, Addison-Wesley, fourth ed., 2011.