Populations, Samples, and Variables

Dr Tom Ilvento
Department of Food and Resource Economics

UNIVERSITY OF DELAWARE

Overview

- We need terms and definitions
  - Populations
  - Samples
  - Measurement
  - Variables
- Types of variables
- How we collect data

What is the population?

- A **Population** is the total number of units involved in the research question. The units are the members (or elements) of the population.
- Populations could be:
  - People
  - Animals
  - Plants
  - Courses
  - Objects
  - Places

A Population is defined by

- Purpose of the study
- The units and elements involved
- Geographic coverage
- Time frame

Example of a population

- What if I was involved in the tourism industry
- And I was interested in understanding the affect of current gas prices on travel in the Mid-Atlantic states, I might define the population as:
- **All households in in the Mid-Atlantic states (DE, MD, PA, NJ, NY) in the Fall of 2008**

Parameters versus Statistics

- A numerical summary measurement describing a population is called a **parameter**
- A numerical summary measurement describing a sample is called a **statistic**.
- Parameters are often represented by Greek characters
  - μ (mu) is often used to represent the population mean
  - Whereas $\overline{X}$ is used for the sample mean

Samples

- When we collect data on all elements in a population, we take a **census**
- However, sometimes it is difficult to get information on the entire population
- So we take a sample of the population
- A **sample** is a subset of the units or elements of a population

Why sample?

- It saves time
- Money
- Other resources (computation time, good will)
- It may actually be impossible to collect information on everyone
  - The satisfaction level of every customer
  - Every part that comes off an assembly line
  - Every black bear in a national park
  - Every person in China

Reoccurring Census Debate

- Every 10 years we take a census

- It is mandated in the Constitution

- However, the Census Bureau knows that it doesn't get a complete count - some groups are difficult to contact

- So, the Census Bureau wants to take a really good sample to estimate the undercount, and then adjust the counts to reflect the missing people

- **The Supreme Court ruled they cannot rely on a sample to make adjustments**

Samples

- Samples are also defined in the terms we used for populations

  - purpose of the research,

  - the units and elements involved,

  - the geographic coverage, and

  - the time frame

Samples

- A valuable property of a sample is that is **representative** of the population.

- By representative we mean the sample characteristics resemble those possessed by the population

- We cannot guarantee the sample will be representative, but if we select it on a random basis we have a probability that it will be representative

- Inferential statistics requires a sample be taken on a random basis

What is a random sample?

- A **random sample** is when each element or unit has the same or nearly the same chance of being selected

- Example

  - If we select a random sample of 100 from a population of 100,000,

  - Each unit has a 100/100,000 or

  - 1/1000th chance of being selected

More Terms

- A **variable** is a characteristic of an individual unit of the population
- Examples are
  - Age of a person in a survey
  - Height of a plant in an experiment
  - Weight of a package in a manufacturing process
- To be a variable the characteristic must vary!!
- It can't all be the same. Otherwise, it's a constant.

Typical Data Structure and Examples of Variables

Subject	Age	Weight	Height	Gender
1	24	162	72"	Male
2	36	180	66"	Male
3	18	136	68"	Male
4	30	162	74"	Male
5	31	186	70"	Male
6	27	144	70"	Male

Measurement

- Measurement is the process of assigning a number or characteristic to variables of the individual units
- Some measurement seems relatively straight-forward
  - years of age
  - dollars of income
  - cholesterol counts
  - parts per million of a chemical
  - gender

Measurement

- Some concepts are no so easy to measure
  - Attitudes
  - Family Type
  - Emotions
  - Intelligence

- What is LOVE?

Measurement

- The process of measurement is often complex – don't take it for granted
- Consider **unemployment**
  - What do we mean by being employed?
  - Are we simply employed or not employed?
- Measurement always comes with some error
- And perhaps Bias

  BIAS is a systematic distortion of the results due to poor design or measurement

17

Measurement

- With measurement we must also deal with
  - **Validity** – are we measuring what we think we are measuring
  - **Reliability** – is the measuring device consistent

18

Level of Measurement

- There are many ways to express the types of variables you might encounter
- One method is to break it down to
  - Quantitative
  - Qualitative
- A more refined method is
  - **Continuous**
  - **Ordinal**
  - **Nominal**

19

Nominal Level Variables

- **Nominal** (or categorical) – no implied order or superiority
  - Male and Female
  - Treatment and Control groups
  - Religious affiliation
- With nominal data I can represent a category with alphanumeric characters or a number, but the number does not imply order or size
  - Code as Yes or No
  - 1=Yes and 0=No

20

Ordinal Level Variables

- **Ordinal** – an implied order or rank, but the distance between units is not well specified

  - Ranking

  - Grades

  - Strongly agree to Strongly disagree

  - A scale from 1 to 10

- I can also use characters or numbers for ordinal level data

Continuous Level Variables

- **Continuous** (combination of interval and ratio) – data that is measured on an infinite scale where we can say something about the magnitude between numbers

  - Temperature

  - Age

  - Income

  - Years of School

Why is the level of measurement important?

- Statistical techniques are predicated on certain levels of measurement.

- Each technique/formula assumes a certain level is used.

- Misusing a statistical technique on a variable can lead to results that are biased or misleading.

JMP and Variable Levels

- JMP will assign variables to one of three categories

  - Categorical

  - Ordinal

  - Continuous

- The assignment will influence what you can do with the variable in terms of an analysis

- You can change the assignment by clicking on the variable name

From where might we get data?

- **Data from a published source** – also known as existing data. Someone else collected it and makes it available to you

  - Census of Population

  - Current Population Survey

  - Company data on customers

- **Caution** – data decisions are out of your control

25

From where might we get data?

- **A Designed Experiment** where the researcher has strict control over the units (people, objects and events).

  - Treatment and Control Groups

  - Randomization of subjects to treatments

- Subjects are assigned randomly or through a controlled process to match subjects

- An experimental design allows you to control more factors and to extract more information from the data

26

From where might we get data?

- **Surveys** are where a researcher samples a group of people, asks a set of questions, and records the answers

  - **Interview**

    - Face-to-Face

    - Telephone

  - **Self-Administered**

    - Mail

    - Internet

- Social Surveys are extremely popular today

27

From where might we get data?

- **Observational Studies** are when the researcher observes the units in their natural setting and records the variables of interest.

  - Focus Groups

  - Studies of consumer behavior

  - Wildlife studies in natural habitat

- Observational Studies must deal with a number of methodological issues

28

Summary

- Things we talked about

  - Population versus a sample

  - The notion of randomness

  - What is a variable?

  - Levels of Measurement

  - Ways in which we collect data

- All of these issues are important in critically
  looking at statistical results

29