

# Business Statistics Chapter Notes

## Chapter 2: Data

### Statistics

- **Descriptive Statistics** (coping with lots of numbers)
  - o Draw a picture (graph, charts, etc)
  - o Calculate a few numbers which summarize the data (mean, median, percentile)
- **Inferential Statistics**
  - o Making decisions and predictions about a population
  - o Generalize facts from a sample to an entire population
  - o Sampling is cheaper, faster, and more accurate
    - Random Sampling
    - Collect the five W's, Who, What, Where, When, and Why
    - Important: if measured in a different time/location, and tell the difference between insight/nonsense

### What is Data?

- Data values or observations are information collected regarding some subject
  - o Numbers, names, etc.
  - o Identifies the Who and What
- Useless without their context
- Data tables are used to help keep the information organized
  - o Rows correspond to individual cases about Whom we record some characteristics
    - Respondents: individual's who answer a survey
    - Subjects or Participants: people in an experiments
    - Experimental Units: animals, plants, websites, or other inanimate objects

### Variables

- Each question measures some aspect of you
- Variable: the aspect / characteristic that differs from subject to subject, individual to individual.
  - o Age, sex, major
- **Data**: the value of the variables
  - o 20, Male, English
- Identifies *What* has been measured, usually shown on the columns

### Variable Types

- **Categorical Variable**: a variable that names categories (whether with words or numeral's)
  - o Descriptive responses to questions
    - What kind of advertising do you use?
  - o Yes or No responses
- **Quantitative Variable**: measured numerical values with units and variable that explains the quantity
  - o Units: yen, cubits, carats, angstroms, nanoseconds, miles per hour, degrees Celsius

### Qualitative or Categorical Variables

- **Ordinal Variables**: categories that have a natural ordering
  - o Numbers
  - o Class such as Freshman, Sophomore, Junior, Senior as (1, 2, 3, 4)
  - o Grades A, B, C, D, F (GPA)
  - o Preference Strongly Agree, Agree, Disagree, Strongly Disagree
- **Nominal Variables**: categories that have no natural ordering
  - o Major and Eye Color

### Two Types of Variables

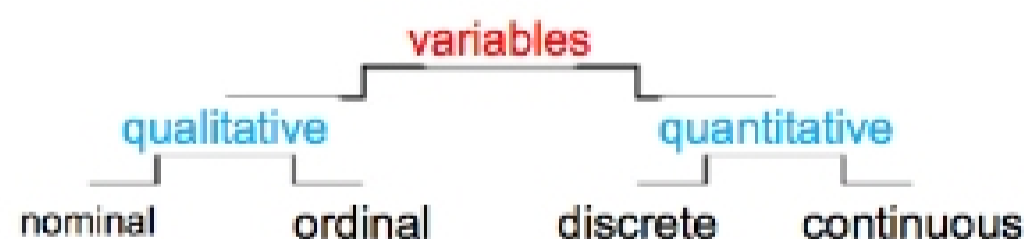
- **Quantitative or Numerical Variables**
  - o Numbers, Measurements, Age, Height, Miles traveled
- **Qualitative or Categorical Variables**
  - o Classifying each observation
  - o Sex, Year in school, major
- **Discrete Variables**: natural gap between the values
  - o # of children / credit cards
  - o Whole #'s
- **Continuous Variables**: values can be arbitrarily close together
  - o Weight, Height, Age
  - o Including decimal's

### Variable Types

- **Counts**: natural way to summarize a categorical / distribution of the variable or data values
  - o Summary of the frequency of cases in a category
  - o Values of a variable whose units are "number of something"
- **Identifiers**: unique identifier assigned to each individual or item in a group
  - o Social Security number, student ID numbers, tracking numbers

### Other Data Types

- Categorical Variables
  - o Nominal Variables
  - o Ordinal Values



### Examples

1. Appraisal of a company's inventory level (excellent, good, fair, poor) = Qualitative, Ordinal
2. Mode of transportation to work (Automobile, bicycle, bus, subway, walk) = Qualitative, Nominal
3. Speed of a Vehicle = Quantitative, Continuous
4. # of persons in each family = Quantitative, Discrete

**Interval Data**: no meaningful zero point

- Can't multiply or divide but the difference between two values is meaningful

- o Temperature

**Ratio Data:** meaningful zero point; can multiply and divide

- o Income, weight, height

**Time Series Data:** ordered data values over time

- o Measured at regular intervals over time
- o Example: determining total costs each month of a year

**Cross-Sectional Data:** data values observed at a single point in

- o Several variables are all measured at the same time point
- o Determining sales revenue, # of customers, and expenses for the last month of business

	Sales (in \$10000's)			
	2003	2004	2005	2006
New York	435	460	475	490
Dallas	320	345	375	395
Seattle	405	390	410	395
Orlando	260	270	285	280

Cross Sectional Data

Time Series Data

time

### Chapter 3: Surveys and Sampling, Key Ideas

#### - 1. Examine a Part of the Whole

- o Draw a Sample
- o **Goal:** learn about an entire population of individual's, but examining all of them is not feasible
- o **Sample:** study a smaller, random group chosen from the population
- o **Biased:** sample doesn't represent the population through over or underemphasize some characteristics

#### - Sources of Bias

- o **Selection B:** difference between population of interest and effective population
- o **Non-Response B:** subjects don't answer or skip questions
- o **Response B:** subjects lie or interviewer effect

#### - Telephone Poll B.

- o Selection B: cell phones or multiple phones
- o Non-response B: answering machines
- o Response B.

#### - 2. Randomize

- o Protect against factors that you know are in the data or unaware of
- o Gets rid of biases
- o Makes sure that on the average the sample looks like the rest of the population
  - Sampling Error: Sample to sample differences

#### - 3. Sample Size Matters

- o Size of the sample matters, not the size of the population
  - Exception: if population is small enough and the sample is more than 10% of the whole population, then it may matter