



Chapter 1 Data and Statistics



General Concept of Statistics

- Techniques by which **data** is processed into **information**
- **Data**
 - observations, facts, measurements in **raw** form
- **Information**
 - data **refined** into a form that is meaningful to a decision maker
- Statistics involves collecting, organizing, presenting, analyzing and interpreting data

2



Applications of Statistics

- Statistical procedures are **tools** which facilitate decision making and problem solving
- Statistics are used in virtually every facet of business and modern life
- **Statistical Literacy**
 - even if you won't be a **preparer** of statistical results, you'll be a **consumer** of them

3

Variables: Quantitative vs Qualitative

- Quantitative
 - The variable is inherently **numeric**
 - egs: checking account balance, minutes remaining in class
- Qualitative
 - The variable being studied is **non-numeric**
 - text codes or arbitrary numeric codes
 - egs: gender, cellular carrier, class rank (Fr, So, Jr, Sr)

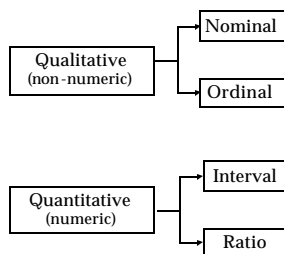
4

Variables: Scales of Measurement

- 4 Scales (Levels) of Measurement:
 - Nominal
 - Ordinal (Rank)
 - Interval
 - Ratio
- Each statistical technique requires a particular **level** of data
 - eg: can't compute an average for nominal-level data

5

Variables



6



Nominal Measurement

- Data values that represent classification **categories**
- Examples
 - Gender (codes of M, F or 0, 1)
 - Blood type (A, B, AB, O, etc.)
- Operations
 - Can compare for equality (=)
 - No order or ranking sequence (<, >)
 - Arithmetic operations are **not** meaningful

7



Ordinal Measurement

- Data values that represent classification **categories** that can be **ranked** with a meaningful order
- Examples
 - Class rank (Fr, So, Jr, Sr)
 - Survey response scale (SA, A, D, SD or 1, 2, 3, 4)
- Operations
 - Can compare for equality (=)
 - Can compare for order or ranking sequence (<, >)
 - Arithmetic operations are **not** meaningful

8



Interval Measurement

- Measurements on a **numerical** scale in which the value zero is **arbitrary**
 - There is no natural zero point
- Examples
 - Fahrenheit temps (30° to 35° & 80° to 85° are the same 5° difference)
 - GMAT scores (possible scores range 200-800)
- Operations
 - Can make comparisons (=, <, >)
 - Differences between interval-level values *are* meaningful
 - Ratios are *not* meaningful (since 0 is arbitrary)

9



Ratio Measurement

- Measurements on a **numerical** scale in which zero is a **meaningful value**
 - The value 0 means **none**
- Examples
 - Weight
 - Bank balance
 - Distances
- Operations
 - Can make comparisons ($=$, $<$, $>$)
 - Differences between ratio-level values **are** meaningful
 - Ratios between ratio-level values **are** meaningful

10



Which Scale (Level)?

- How would you like your steak cooked?
- What year is it?
- What is your eye color?
- How many minutes does it take you to drive to school?
- How many bedrooms where you live?

11



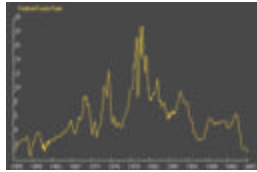
Continuous vs Discrete (Ch 5)

- Continuous
 - Result from measurement
 - Can take on any value (including fractions) along the relevant portion of the scale of measurement
 - Example: weight
- Discrete
 - Result from counting
 - Can take on only a finite number of (usually integer) values along the relevant portion of the scale
 - Example: number of students absent

12

Cross Sectional and Time Series Data

- Cross Sectional
 - Observations are collected at the same point in time
 - Eg: exam score data set
- Time Series
 - Observations are collected over several time periods
 - Monitor important variables over time
 - Eg: monthly unemployment, annual sales, daily absences



13

Data Sources

- Existing Sources
 - Obtain from operational databases, government agencies, published studies, industry trade associations, web sources
- Statistical Studies
 - Endure cost & time associated with obtaining own measurements when data is not readily available from existing sources
 - Experimental
 - elements are under the control of the researcher
 - elements can be assigned to different treatment groups
 - eg: new drug safety and effectiveness
 - Observational
 - elements not under the control of the researcher
 - eg: surveys, polls
 - eg: exam score data set

14

Main Branches of Statistics

- Descriptive
 - Chapters 2 and 3
 - Graphical & numerical methods of summarizing and describing data
- Inferential
 - Chapters 7+
 - Deriving conclusions about a **population** using **sample data**
 - Population characteristics are estimated by analyzing sample data
 - Confidence intervals, hypothesis testing, relationships among variables
- Probability theory
 - Chapters 4-6
 - Used to measure the risk of sampling error & therefore deriving wrong conclusion about the population

15



Some Basic Concepts

- Population (universe)
 - The entire membership of some defined group of interest
 - Census (nib)
 - Parameter
- Sample
 - A subset of a population selected for study
 - Statistic
- Summary
 - Population parameters are estimated by sample statistics
- Advantages and Disadvantages of Sampling

16



Data Acquisition Errors

- Error = (true value – estimated value)
= (parameter – statistic)
- Main causes:
 - recording error/typo
 - sampling error
 - representative vs. non-representative samples
 - response bias
 - question wording

17
