# Chapter 1
# Introduction to Statistics

Slide 1

A) Overview

B) Types of Data

C) Critical Thinking

D) Random Sampling



## Definitions

Slide 2

❖ **Statistics**

**a collection of methods for planning experiments, obtaining data, and then then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data.**



## Definitions

Slide 3

❖ **Data**

**observations (such as measurements, genders, survey responses) that have been collected.**



## Definitions

Slide 4

❖ **Population**

**the complete collection of all elements (scores, people, measurements, and so on) to be studied. The collection is complete in the sense that it includes all subjects to be studied.**



## Definitions

Slide 5

❖ **Census**

**the collection of data from every member of the population.**

❖ **Sample**

**a sub-collection of elements drawn from a population.**



## Key Concepts

Slide 6

❖ **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.**

# Definitions

Slide 7

❖ **Parameter**

a numerical measurement describing some characteristic of a **population**

population

⇕

parameter



# Definitions

Slide 8

❖ **Statistic**

a numerical measurement describing some characteristic of a **sample**.

sample

⇕

statistic



# Definitions

Slide 9

❖ **Quantitative data**

numbers representing counts or measurements.

Example: weights of supermodels.



# Working with Quantitative Data

Slide 10

**Quantitative data can further be distinguished between discrete and continuous types.**



# Definitions

Slide 11

❖ **Discrete**

data result when the number of possible values is either a finite number or a 'countable' number of possible values.

0, 1, 2, 3, . . .

Example: The number of eggs that hens lay.

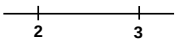


# Definitions

Slide 12

❖ **Continuous**

(numerical) 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.

# Definitions


❖ **Qualitative (or categorical or attribute) data**

**can be separated into different categories that are distinguished by some nonnumeric characteristics.**

**Example: genders (male/female) of professional athletes.**



# Levels of Measurement


**Another way to classify data is to use use levels of measurement. Four of these levels are discussed in the following slides.**



# Definitions


❖ **nominal level of measurement**

**characterized by data that consist of names, labels, or categories only. The data cannot be arranged in an ordering scheme (such as low to high)**

**Example: survey responses yes, no, undecided**



# Definitions


❖ **ordinal level of measurement**

**involves data that may be arranged in some order, but differences between data values either cannot be determined or are meaningless**

**Example: Course grades A, B, C, D, or F**



# Definitions


❖ **interval level of measurement**

**like the ordinal level, with the additional property that the difference between any two data values is meaningful. However, there is no natural zero starting point (where *none* of the quantity is present)**

**Example: Years 1000, 2000, 1776, and 1492**



# Definitions


❖ **ratio level of measurement**

**the interval level modified to include the natural zero starting point (where zero indicates that *none* of the quantity is present). For values at this level, differences and ratios are meaningful.**

**Example: Prices of college textbooks ($0 represents no cost)**

## Summary - Levels of Measurement

Slide 19

- **Nominal** - categories only
- **Ordinal** - categories with some order
- **Interval** - differences but no natural starting point
- **Ratio** - differences and a natural starting point



## Success in Statistics

Slide 20

- Success in the introductory statistics course typically requires more common sense than mathematical expertise.
- This section is designed to illustrate how common sense is used when we think critically about data and statistics.



## Misuses of Statistics

Slide 21

- Bad Samples
- Small Samples
- Misleading Graphs
- Pictographs
- Distorted Percentages
- Loaded Questions
- Order of Questions
- Refusals
- Correlation & Causality
- Self Interest Study
- Precise Numbers
- Partial Pictures
- Deliberate Distortions



## Definitions

Slide 22

- Voluntary response sample
  (or self-selected survey)

  one in which the respondents themselves decide whether to be included.

  In this case, valid conclusions can be made only about the specific group of people who agree to participate.



Slide 23

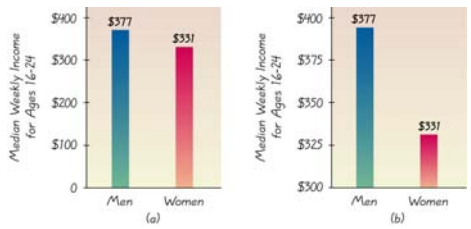



**Figure 1-1**



Slide 24

**To correctly interpret a graph, we should analyze the numerical information given in the graph instead of being mislead by its general shape.**

Slide 25

**Double the length, width, and height of a cube, and the volume increases by a factor of eight**

Figure 1-2



Slide 26

**97% yes: "Should the President have the line item veto to eliminate waste?"**

**57% yes: "Should the President have the line item veto, or not?"**



# Definitions

Slide 27

❖ **Random Sample**

members of the population are selected in such a way that each individual member has an equal chance of being selected

❖ **Simple Random Sample (of size *n*)**

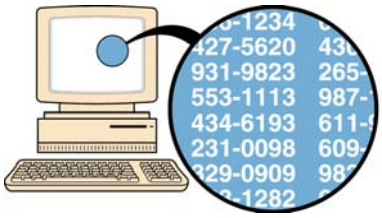
subjects selected in such a way that every possible sample of the same size ***n*** has the same chance of being chosen



## Random Sampling

Slide 28

**selection so that each has an equal chance of being selected**






## Systematic Sampling

Slide 29

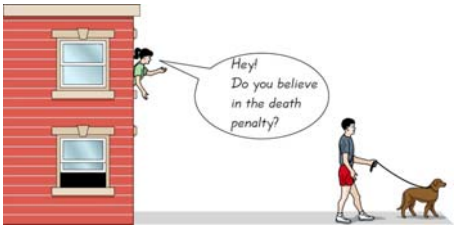
**Select some starting point and then select every K th element in the population**



## Convenience Sampling

Slide 30

**use results that are easy to get**



# Stratified Sampling

Slide 31

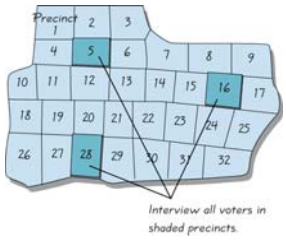
**subdivide the population into at least two different subgroups that share the same characteristics, then draw a sample from each subgroup (or stratum)**




# Cluster Sampling

Slide 32

**divide the population into sections (or clusters); randomly select some of those clusters; choose all members from selected clusters**






# Definitions

Slide 33

- **Sampling Error**
  the difference between a sample result and the true population result; such an error results from chance sample fluctuations

- **Nonsampling Error**
  sample data that are incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective instrument, or copying the data incorrectly)