Types of Data, Mean and Median

Section 1.1 & 1.2

Cathy Poliak, Ph.D.
cathy@math.uh.edu

Department of Mathematics
University of Houston

January 19, 2016

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What is “Data”?

- The facts and figures collected,analyzed, and summarized for presentation and interpretation.
- Amount of your last purchase at a grocery store.
- The number of times that you access a certain website.
- Your name.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Data of basketball shoes

From footlocker.com the following is an excerpt form the list of data of 91 different pairs of men's basketball shoes. This data set is named **basketball shoes**.

Name	Brand	Price
adiPower Howard 2	Adidas	75
adiZero Crazy Light	Adidas	90
adiZero Crazy Light 2	Adidas	140
⋮		
1 Flight	Nike Jordan	100
1 Flight Low	Nike Jordan	95
⋮		
Air Max CB34	Nike	110
Air Max Dominate	Nike	75
⋮		

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Break down of the data set

- **All pairs of men's basketball shoes** are defined to be the *cases*. The **cases** are the objects described by a set of data. Not necessarily people.
- The **name of the shoe** is considered to be the *label*. A **label** is a special variable used in some data sets to distinguish the different cases.
- The **brand of the shoe** and **price** are the *variables* for this data set. A **variable** is any characteristic of an individual or object. A variable can take on different **values** for different individuals or objects.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Types of data

- **Population Data** is everything or everyone we want information about. It is a set of data that consists of all possible values pertaining to a certain set of observations or an investigation.

- **Sample Data** is a subset of the population that we have information from. It is just a small section of the population taken for the purpose of investigation.

Examples of Types of Data

Identify the population and the sample for each of the following:

- University of Houston is interested in how many students buy used books as opposed to new ones. They randomly choose 100 students at the student center to interview
  - Population -
  - Sample -
- An elementary school is creating a new lunch menu. They send questionnaires to students with last names that begin with the letters M through R.
  - Population -
  - Sample -

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Two Types of Variables

Go back to the example of the basketball shoes. We have two variables, brand of the shoe and price of the shoe.

- The variable **brand of the shoe** is a *categorical variable*. **Categorical variables** place a case into one of several groups or categories.
- The variable **price** is a **quantitative variable**. **Quantitative Variables** take numerical values for which arithmetic operations such as adding and averaging make sense.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Two Types of Quantitative Variables

Quantitative variables can be classified as either **discrete** or **continuous**.

- Discrete quantitative variables - a countable set of values.
- Continuous quantitative variables - data that can take on any values within some interval.
- What type of quantitative variable is **price**?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples of Variables

Classify the following variables as categorical or quantitative. If quantitative, state whether the variable is discrete or continuous.

- Political preference.

- Number of siblings.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples of Variables Part 2

Classify the following variables as categorical or quantitative. If quantitative, state whether the variable is discrete or continuous.

- Blood type.
- Height of men on a professional basketball team.
- Time it takes to be on hold when calling the IRS at tax time.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Describing Quantitative Variables with Numbers

- Center - mean, median or mode
- Spread - range, interquartile range, variance, or standard deviation
- Location - percentiles or standard scores

Parameters and Statistics

- A **parameter** is a number that describes the **population**. A parameter is a fixed number, but in practice we usually do not know its value.
- A **statistic** is a number that describes a **sample**. The value of a statistic is known when we have taken a sample, but it can change from sample to sample. We often use a statistic to estimate an unknown parameter.
- The purpose of sampling or experimentation is usually to use statistics to make statements about unknown parameters, this is called **statistical inference**.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Notation of Parameters and Statistics

Name	Statistic	Parameter
mean	$\bar{x}$	μ mu
standard deviation	s	σ sigma
correlation	r	ρ rho
regression coefficient	b	β beta
proportion	$\hat{p}$	p

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

A carload lot of ball bearings has a mean diameter of **2.503** centimeters. This is within the specifications for acceptance of the lot by the purchaser. The inspector happens to inspect 100 bearings from the lot with a mean diameter of **2.515** centimeters. This is outside the specified limits, so the lot is mistakenly rejected. Is each of the bold numbers a parameter or a statistic?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Presidential Approval Rating

In a survey conducted between September 20, 2012 and September 22, 2012 by Gallup.com, 51% of Americans approved of how Obama is doing as President. Gallup tracks daily the percentage of Americans who approve or disapprove of the job Barack Obama is doing as president. Daily results are based on telephone interviews with approximately 1,500 national adults; Margin of error is $\pm$ 3 percentage points.
Is this 51% a statistic or parameter?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example: Nike shoes

We want to know some information about the variable **price**.

Name	Price
LeBron 9 PS Elite	250
LeBron 9 Limited iD	215
Nike Zoom Kobe VII System	140
Nike Kobe VII System Low iD	185
LeBron 9 iD	215
Nike Kobe VII System Mid iD	185
Nike Zoom KD IV iD	140
Lebron 9 Low	150
Nike Zoom Soldier VI	120
Nike Hyperdunk	250
Nike Lunar Hyperdunk 2012	140
Nike Lunar Hyperdunk iD	290
Nike Hyperfuse 2012	110
Nike Zoom Soldier VI	120
Air Max Hyperaggressor	100

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Measuring center: The mean

- Most common measure of center.
- Arithmetic average.
- To calculate the mean of a set of observations $x_1, x_2, \ldots, x_n$, add their values and divide by the number of observations ***n***.
- Denoted: $\bar{x}$ called x-bar if the data is from a sample, μ, called "mu" if the data is from the entire population.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{n} x_i$$

- Where n is the size of the sample and N is the size of the population.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sample Mean of Price

$$\bar{x} = \frac{1}{15} \times (250 + 215 + 140 + 185 + 215 + 185 + 140 + 150 + 120 + 250 + 140 + 290 + 110 + 120 + 100)$$

$$\bar{x} = \frac{2610}{15}$$

$$\bar{x} = 174$$

The sample mean price of these men's Nike basketball shoes is \$174.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Measuring center: The Median

The **median** M is the midpoint of a data set such that half of the observations are smaller and the other half are larger.

1. Arrange all observations in order of size, from smallest to largest.
2. Find the middle value of the arranged observations by counting $(n+1)/2$ from the bottom of the list.
   - If the number of observations $\boldsymbol{n}$ is odd, the median M is the the center observation in the ordered list.
   - If the number of observations $\boldsymbol{n}$ is even, the median M is the mean of the two center observation in the ordered list.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Median of Basketball Shoe Prices

1. Arrange the prices in order from lowest to highest.

100	110	120	120	140	140	140	150
185	185	215	215	250	250	290	

2. The middle value is in the $\frac{15+1}{2} = 8^{\text{th}}$ place.
3. The median is \$150.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Measuring Center: The Mode

- The **mode** of a data set is the numerical value that appears the most frequently.
- The data set can have one mode, two or more modes.
- A data set may not have any mode.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Mode of Basketball Shoe Prices

- The following are the prices of the basketball shoes arranged in order:

  100 110 120 120 140 140 140 150
  185 185 215 215 250 250 290

- There are three 140 this is the most frequent value. Thus the mode for the price of basketball shoes is $140.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example: Speaking Age

Twelve babies spoke for the first time at the following ages (in months):

8 9 10 11 12 13 15 15 18 20 20 26

- What is the mean of the data?
- What is the median of the data?
- What is the mode of the data?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example: Weights of Steers

Here are the weights (in pounds) of 20 steers on an experimental feed diet:

174	142	131	145	175	150	176	151	110	162
133	163	135	178	178	154	166	146	156	167

- What is the mean of the data?

- What is the median of the data?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example: Test Scores

The test scores of a class of 20 students have a mean of 71.6 and the test scores of another class of 14 students have a mean of 78.4. Find the mean of the combined group.

Example: Conclusions

- A businesswoman calculates that the median cost of the five business trips that she took in a month is $600 and concludes that the total cost must have been $3000.
- Explain why the conclusion drawn is not valid.

Finding the Mean and Median in R

This is the R code to input data to get mean and median

```
> price<-c(250,215,140,185,215,185,140,150,120,250,140,290,110,120,100)
> mean(price)
[1] 174
> median(price)
[1] 150
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Mean vs. Median

- If the mean and the median are both numbers that describe the center of the values then why do we have different values?
- If the data has values that are **outliers** values that are beyond the range of the others, the mean is going toward these outliers.
- The median is resistant to extreme values (outliers) in the data set.
- The mean is NOT robust against extreme values.
- We will discuss this more with the graphs in section 1.5.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS