STA 1013 : Statistics through Examples

Lecture 7: Measure of Center

Hwiyoung Lee

September 16, 2019

Department of Statistics, Florida State University

Overview

Measure of Center

Measure of Center

Measure of Center

- Mean
- Median
- Mode

Mean

The **mean** is what we most commonly call the average value

$$\text{Mean} = \frac{\text{sum of all values}}{\text{total number of values}} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

- x_i : i th observation
- Example : $1, 2, 2, 4, 5, 10$

$$\frac{1+2+2+4+5+10}{6} = 4$$

Mean

- Most familiar measure of center
- Takes every value into account
- Affected by outliers

Example : Annual Income

Name	Annual Income
Tom	$ 32,000
Larry	$ 36,000
Susan	$ 39,000
Paul	$ 41,000
Marcus	$ 50,000
Randy	$ 57,000
Sandy	$ 60,000
Tim	$ 75,000
Pam	$ 80,000
Kim	$ 95,000

Mean ?

Example : Annual Income

Bill Gates moves to town

Name	Annual Income
Tom	$ 32,000
Larry	$ 36,000
Susan	$ 39,000
Paul	$ 41,000
Marcus	$ 50,000
Randy	$ 57,000
Sandy	$ 60,000
Tim	$ 75,000
Pam	$ 80,000
Kim	$ 95,000
Bill Gates	$ 5,000,000,000

Mean ?

Outlier

Outlier

a value that is much higher or much lower than almost all other values.

- Bill Gates's annual income \$ 5,000,000,000 would be an outlier
- An outlier can pull the mean significantly upward (or downward)
- Making the mean unrepresentative of the data set as a whole

Example : Outlier

Example

Imagine that the 5 graduating seniors on a college basketball team receive the following first-year contract offers to play in the NBA (zero indicates that the player did not receive a contract offer):

0	0	0	0	$ 10,000,000

Table 1: Graduating seniors NBA contract offer

- Is it therefore fair to say that the average senior on this basketball team received a $2 million contract offer?
- Not really. The single player receiving the large offer makes the mean much larger than it would be otherwise.

Median

Median : the middle value of the ordered data

- Half of the observations are larger and half are smaller than the Median.
- How to get ?
    1. Sort the data ascending order : the lowest value to the highest value.
    2. The Median is the $(n+1)/2$ th observation
    3. If n is odd, the median is the middle observation of the ordered array.
    4. If n is even, it is midway between the two central observations.
- Median is robust to ourliers

Example (n = odd)

Example : 2, 4, 1, 10, 8, 9, 5

- Sort : 1, 2, 4, 5, 8, 9, 10
- Median : $(7+1)/2$ th observation
  - 4 th observation : 5
- Q: What happens to the median if we change the 10 to 1,000?

- Mean :
- Q: What happens to the mean if we change the 10 to 1,000?

Example (n = even)

Example : 5, 11, 1, 13, 6, 9, 8, 3

- Sort : 1, 3, 5, 6, 8, 9, 11, 13
- Median : $(8+1)/2$ th observation
  - 4.5 th observation : average of 4 th and 5 th observation
- Q: What happens to the median if we change the 1 to -1,000?

- Mean :
- Q: What happens to the mean if we change the 1 to -1,000?

Example : Annual Income

Bill Gates moves to town

Name	Annual Income
Tom	$ 32,000
Larry	$ 36,000
Susan	$ 39,000
Paul	$ 41,000
Marcus	$ 50,000
Randy	$ 57,000
Sandy	$ 60,000
Tim	$ 75,000
Pam	$ 80,000
Kim	$ 95,000
Bill Gates	$ 5,000,000,000

Median ?

Midrange

Midrange

- The value that midway between the maximum and minimum in the data set

$$\text{Midrange} = \frac{\text{Max} + \text{Min}}{2}$$

- Easy to calculate
- Example : 1, 2, 4, 5, 7, 9, 10, 11
  - Midrange :
  - Median :

Midrange

- Because the midrange uses only the maximum and minimum values (ignore all other values), it is very **sensitive to those extremes**
- Example : 0, 5, 5, 5, 6, 7, 8, 9, 100
    - Midrange :
    - Median :
- Not popular (Rarely used)

Mode

Mode : the value of the data that occurs with the greatest frequency.

Example : 1, 2, 3, 3, 3, 4, 5

- Mean :

- Median :

- Mode :

Mode for skewed data

Example : 5, 5, 5, 6, 6, 7, 8, 9, 10, 11, 100 , 500, 1,000,000

- Mean :
- Median :
- Mode :

Mode for Right skewed data

Example : 5, 5, 5, 6, 6, 7, 8, 9, 10, 11, 100 , 500, 1,000,000

- Mean : 76,974.77

- Median : 8

- Mode : 5

Mode $<$ Median $<$ Mean

Mode for skewed data

Example : -1,000,000, -500, -100, 6, 6, 7, 7, 7, 8, 8, 8, 8, 10

- Mean :
- Median :
- Mode :

Mode for Left skewed data

Example : -1,000,000, -500, -100, 6, 6, 7, 7, 7, 8, 8, 8, 8, 10

- Mean :-7,732.692
- Median : 7
- Mode : 8

$$\textbf{Mode} > \textbf{Median} > \textbf{Mean}$$

Mode

Mean, Median always exist and are always unique

1. The mode **may not exist**
   - Example : 1, 2, 3, 4, 5, 6, 7, 8, 9
2. The mode **may not be unique**
   - Example : 1, 2, 2, 2, 5, 6, 7, 7, 7, 8, 8, 9, 10, 10, 10

Mode for continuous data

Example : Weight data

195.6	200.4	165.6	165.3	191.7	169.3	153.2
189.5	170.4	149.3	185.3	150.3	179.6	160.3
198.5	163.2	166.3	197.3	201.3	168.2	198.4

- Mean : 177.0952

149.3	150.3	153.2	160.3	163.2	165.3	165.6
166.3	168.2	169.3	**170.4**	179.6	185.3	189.5
191.7	195.6	197.3	198.4	198.5	200.4	201.3

Table 2: Sorted weight data

- Median : 170.4

For continuous data, Mode usually does not exist !!

Mode for binned data

Binned weight data

Weight	Count
140 $\sim$ 149.9	1
150 $\sim$ 159.9	2
160 $\sim$ 169.9	**7**
170 $\sim$ 179.9	2
180 $\sim$ 189.9	2
190 $\sim$ 199.9	5
200 $\sim$ 209.9	2

- **Mode** : the bin with the highest frequency (160 $\sim$ 169.9)
- Binned data will be used for the Histogram

Weighted Mean

Motivation

Suppose two differnt group : A, B

Group	n	Mean
A	10	20
B	30	40

- The overall mean of 40 people

$$\frac{20+40}{2} = 30 \ ?$$

Totally wrong !!

The sizes of the two different groups must be taken into account

Motivation

Recall the mean formula :

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\text{Total sum of all data values}}{\text{The number of all data values}}$$

Since,

- Group A

  Total sum : 10 × 20 = 200 ($\because \frac{\text{Total sum of A}}{10} = 20$)

- Group B

  Total sum : 30 × 40 = 1,200 ($\because \frac{\text{Total sum of B}}{30} = 40$)

Thus,

$$\text{Mean} = \frac{10 \times 20 + 30 \times 40}{40} = \frac{200 + 1,200}{40} = 35$$

Motivation

State	n	Mean
Alabama	n_1	$\bar{x}_1$
Alaska	n_2	$\bar{x}_2$
$\vdots$	$\vdots$	$\vdots$
Wisconsin	n_{49}	$\bar{x}_{49}$
Wyoming	n_{50}	$\bar{x}_{50}$

Don't do

$$\bar{x}_{usa} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_{49} + \bar{x}_{50}}{50} \qquad \text{Wrong !!}$$

The sizes of the 50 different states must be taken into account

$$\bar{x}_{usa} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_{49}\bar{x}_{49} + n_{50}\bar{x}_{50}}{n_1 + n_2 + \cdots + n_{49} + n_{50}}$$

Weighted Mean

A weighted mean accounts for variations in the relative importance of data values.

$$\text{Weighted mean} = \frac{\text{Sum of (each data value} \times \text{its weight)}}{\text{sum of all weights}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

- Each data value is assigned a weight (w_i)
- Weighted means are appropriate whenever the data values vary in their degree of importance

Example : Grade

Grade

Suppose your course grade is based on four quizzes and one final exam. Each quiz counts as 15% of your final grade, and the final counts as 40%. Your quiz scores are 75, 80, 84, and 88, and your final exam score is 96. What is your overall score?

solution :

$$\frac{15 \times 75 + 15 \times 80 + 15 \times 84 + 15 \times 88 + 40 \times 96}{15 + 15 + 15 + 15 + 40} = \frac{8745}{100} = 87.45$$

Exercise : Grade

Grade

Suppose your course grade is based on four quizzes and one final exam. Your lowest quiz grade will be dropped, and the other three will equally worth 20% of your grade, and the final counts as 40%. Your quiz scores are 100, 70, 64, and 88, and your final exam score is 96. What is your overall score?

solution :

Example : GPA

GPA

Randall has 38 credits with a grade of A, 22 credits with a grade of B, and 7 credits with a grade of C. What is his grade point average (GPA)? Base the GPA on values of 4.0 points for an A, 3.0 points for a B, and 2.0 points for a C.

solution :

$$\frac{38 \times 4 + 22 \times 3 + 7 \times 2}{38 + 22 + 7} = \frac{232}{67} = 3.46$$

Exercise : GPA

GPA

Lee has 30 credits with a grade of A, 15 credits with a grade of B, 5 credits with a grade of C, and 3 credits with F. What is his grade point average (GPA)? Base the GPA on values of 4.0 points for an A, 3.0 points for a B, 2.0 points for a C, and 0 points for a F.

solution :