Measures of Central Tendency - the Median and Mode

Dr Tom Ilvento
Department of Food and Resource Economics

Overview

- We will look at two other measures of Central Tendency
  - **Median** and other positional measures
  - **Mode** – the most frequently occurring value
- Then we will look at how we might compare the mean, median, and mode for data
- I will use the same two data sets - Mariage Rate and Fastest Speed

Median

- The **Median** is the middle value when the measurements are arranged in ascending order.
- It is the value at the **50th percentile**
- It is a **positional measure** because it is based on the middle case in a variable.
- In order to find the median value, we first must:
  1. sort the data in ascending or descending order
  2. Find the middle position
  3. Read the value at the middle position

Finding the Median

- First Sort the data
- Next identify the median position in the data, and is the sample size odd or even?
- n= 51
- In our case the data are odd, so there is an exact middle the 26th observation which is 7.0

Stem & Leaf of Marriage Rate

		Count
4	2 7	2
5	0 5 5 8 9 9	6
6	1 1 1 3 3 4 5 6 6 7 7 8 9 9	14
7	0 0 0 0 3 3 3 4 4 4 7 9	12
8	1 1 2 3 3 4 6 8 9 9	10
9	4 5	2
10	3 5	2
11		0
12	6	1
13		0
14		0
15		0
16		0
17		0
18		0
19		0
20		0
21		0
22	5	1
61	0	1

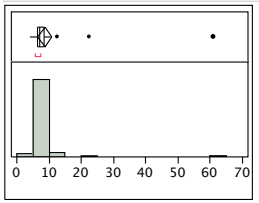
4 2 = 4.2

Finding the Median

- **If n is an odd number**, the median is the (n+1)/2 value in the ordered data.
  - Example: If N=99, then the median value is the value of the (99+1)/2 = 50th case.
  - In our case, n=51 so the median is the (51+1)/2 = 26th observation
- **If n is an even number**, the median value will fall between the n/2th and the (n/2)+1th cases.
  - Example: If N = 100, median value is between the 50th and 51st cases.
  - In this case, we usually take the average of the two values to find the median value.

Result from JMP

2005 Marriage Rate



Quantiles

100.0%	maximum	61.000
99.5%		61.000
97.5%		49.450
90.0%		10.140
75.0%	quartile	8.300
50.0%	median	7.000
25.0%	quartile	6.300
10.0%		5.560
2.5%		4.350
0.5%		4.200
0.0%	minimum	4.200

Moments

Mean	8.6588235
Std Dev	7.92056
Std Err Mean	1.1091002
upper 95% Mean	10.886517
lower 95% Mean	6.4311301
N	51
Sum Wgt	51
Sum	441.6
Variance	62.735271
Skewness	6.0999188
Kurtosis	40.036869
CV	91.473859
N Missing	0

Finding the Median Value

- Once we find the median position, we read the value at that position
- In the case of an even sample size, the average of the n/2th and (n/2)+1th values
- This why we call it a **positional statistic** – it is the value at the position in the data

Properties of the Median

- The median is an intuitive measure of central tendency - the middle.
- Spreadsheets and software packages will now easily calculate the median for us.
- The median has limited inferential properties
- But, it is not as sensitive to outliers and thus is used in data with extreme values
  - Income – a few very wealthy people and many in middle and lower classes
  - Company size – a few large companies and many small ones

Example of the Median's resistance to extreme values

- Marriage rate data for 2005 has two extreme values, 61.0 and 22.5
- If we just remove the highest value, n = 50, which is even!
- Median value is between
  - 50/2 = 25th value
  - and the (50/2)+1 = 26th value
- This is the average of
  - 25th (Nebraska)
  - and 26th (New York) values
  - both of which are 7.0
- **The Median did not change!**

Stem & Leaf of Marriage Rate

		Count
4	2 7	2
5	0 5 5 8 9 9	6
6	1 1 1 3 3 4 5 6 6 7 7 8 9 9	14
7	0 0 0 0 3 3 3 4 4 4 7 9	12
8	1 1 2 3 3 4 6 8 9 9	10
9	4 5	2
10	3 5	2
11		0
12	6	1
13		0
14		0
15		0
16		0
17		0
18		0
19		0
20		0
21		0
22	5	1

Positional Measures

- The median is also referred to as the **50th percentile.**
- There are other ordered measures:
  - Quartiles
    - The first quartile or **Q1** is the 25th percentile
    - **Q3** is the 75th percentile
    - The median is also second quartile or **Q2**
  - Percentiles, deciles, and quintiles are also ordered measures
  - As are the **maximum** and **minimum** values
- **Now we have the five number summary for the Box Plot!!**

JMP Output showing positional measures

2005 Marriage Rate

Quantiles

100.0%	maximum	61.000
99.5%		61.000
97.5%		49.450
90.0%		10.140
75.0%	quartile	8.300
50.0%	median	7.000
25.0%	quartile	6.300
10.0%		5.560
2.5%		4.350
0.5%		4.200
0.0%	minimum	4.200

Moments

Mean	8.6588235
Std Dev	7.92056
Std Err Mean	1.1091002
upper 95% Mean	10.886517
lower 95% Mean	6.4311301
N	51
Sum Wgt	51
Sum	441.6
Variance	62.735271
Skewness	6.0999188
Kurtosis	40.036869
CV	91.473859
N Missing	0

The Mode

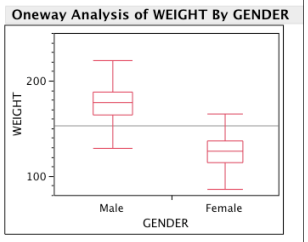
- The mode is the most frequent occurring value in a variable.
- There may not be a single unique value that occurs the most often in continuous level data.
  - In this case the mode is undefined
  - **Software packages may leave it undefined if there is a tie**
- In a qualitative variable, we refer to the Modal Class or Category.
- The mode may make more sense in reference to categorical data.

Bi-Modal Graph of Weight

Stem and Leaf of WEIGHT

Stem	Leaf	Count
24	9	1
23	7	1
22	155	3
21	347	3
20	1122225568999	13
19	0000001122445589	16
18	00111222233333444445555566666778889999	38
17	0011112233344445555566666777777889	36
16	000111123344445566668888	24
15	00111111155556667778889	22
14	000012233445666677788899999	26
13	000111222222333344556666777788889999	36
12	00011122222333344445555666666666677777888999999	46
11	0011222223333444556788899	24
10	0233334444457788	16
9	008999	6
8	6	1

8|6 represents 86



Marriage Rate Example

Stem & Leaf of Marriage Rate

		Count
4	2 7	2
5	0 5 5 8 9 9	6
6	1 1 1 3 3 4 5 6 6 7 7 8 9 9	14
7	0 0 0 0 3 3 3 4 4 4 7 9	12
8	1 1 2 3 3 4 6 8 9 9	10
9	4 5	2
10	3 5	2
11		0
12	6	1
13		0
14		0
15		0
16		0
17		0
18		0
19		0
20		0
21		0
22	5	1
61	0	1

4|2 = 4.2

Mean = 8.66
Median = 7.02
Mode = 7.0

Comparing the Mean, Median, and Mode

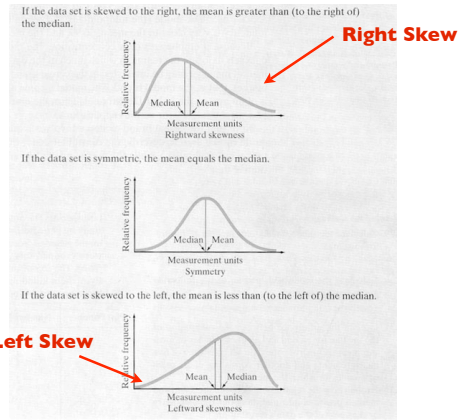
- In a symmetrical, bell shaped curve depicting the distribution of a variable, the mean, median, and mode would the same or very similar values
- The **normal curve** is a very special symmetrical bell shaped curve where the mean, median, and mode are all equal by definition.

Comparing the Mean, Median, and Mode

If the data are skewed to the right, the mean is greater than the median and it is being pulled by extreme values to the right

Mean = Median

If the data are skewed to the left, the mean is less than the median and it is being pulled by extreme values to the left



A Note on Skew

- When we use the term **skew**, we mean a tail in the distribution toward extreme values
- If there is skew right, there are a few extreme values to the right, and most or many of the values are bunched to the left
- If there is skew left, there are a few extreme values to the left, and most or many of the values are bunched to the right

Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

The MPG distribution as a symmetrical, mound-shaped distribution

Mean = 36.994
Median = 37.0
Mode = 37

MPG Stem and Leaf

Stem	Leaf	Count
44	9	1
43		
42	1	1
41	002	3
40	0123557	7
39	00345789	8
38	0122345678	10
37	00001112233445667789 9	21
36	01233445566777888999	20
35	01235667899	11
34	024588	6
33	126899	6
32	5799	4
31	8	1
30	0	1

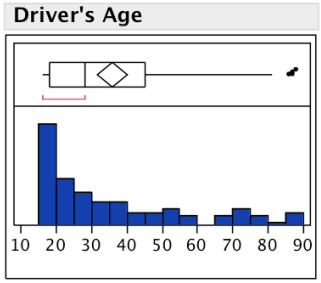
30|0 represents 30.0

Excel Spreadsheet of EPA Data with Excel commands for Central Tendency

MPG				
30.0	**Sum**	=SUM(B5:B104)	3699.40
31.8	**Count**	=COUNT(B5:B104)	100.00
32.5	**Mean**	=AVERAGE(B5:b104)	36.99
32.7	**Minimum**	=MIN(B5:B104)	30.00
32.9	**Maximum**	=MAX(B5:B104)	44.90
32.9	**Median**	=MEDIAN(B5:B104)	37.00
33.1	**Mode**	=MODE(B5:B104)	37.00
33.2	**Quartile**	=QUARTILE(B5:B104,1)	

Let's Look at the Driver's Age data

- The data are skewed right – a few extreme values in the 70s and 80s, but most in the teens and 20s.
- n = 100
- Sum(x) = 3587



Remember

- Sample size n = 100
- Sum(x) = 3,587
- Median position is
  - Between the 100/2 and the 100/2 + 1 positions, i.e.,
  - the 50th and 51st positions
- The Mode is the most frequent number

So if I give you

- n = 100
- Sum(x) = 3587
- You should be able to solve for:
  - **Mean**
  - **Median**
  - **Mode**

Driver's Age Stem and Leaf

Stem	Leaf	Count
8	6778	4
8	1	1
7	667	3
7	03334	5
6	599	3
6		
5	678	3
5	11123	5
4	5555	4
4	0244	4
3	5677888	7
3	0011244	7
2	5677788899	10
2	00011123444444	14
1	66666666777777778888888888889	30

1|6 represents 16

Solution

- **Mean = 3587/100 = 35.87**
- **Median** is the average of the 50th and 51st positions:
- **Median = (28+28)/2 = 28.0**
- **Mode is 18**
- In this case, the mean, median, and mode are all very different
- The data are skewed right and the mean is pulled by the extreme values

Driver's Age Stem and Leaf

Stem	Leaf	Count
8	6778	4
8	1	1
7	667	3
7	03334	5
6	599	3
6		
5	678	3
5	11123	5
4	5555	4
4	0244	4
3	5677888	7
3	0011244	7
2	5677788899	10
2	00011123444444	14
1	66666666777777778888888888889	30

1|6 represents 16

Median falls here

Mode is here

Output from Excel, Descriptives

- In Excel, use Data Analysis
- Descriptives

Age	
Mean	**35.9**
Standard Error	2.1
Median	**28.0**
Mode	**18.0**
Standard Deviation	21.1
Sample Variance	447.1
Kurtosis	0.0
Skewness	1.1
Range	72.0
Minimum	16.0
Maximum	88.0
Sum	3587.0
Count	**100.0**

Summary

- Central tendency of the data – the average or the middle – has appeal of what might be “typical”
- We discussed the
  - Mean
  - Median
  - Mode
- The shape of the distribution can affect how much these measures agree as a measure of center