Research Experience for Teachers: Mitigating Natural Disasters

Lesson #2 Center and Spread

Subject Area	Mathematics
Grade Level	9-12
Prior knowledge	None
Time required	2 class periods

Summary

Descriptive statistics requires the analyst to select the display and summary statistics that are appropriate for the data. In this lesson students will calculate summary statistics to describe a set of numbers. Students will learn that a single measure of center can be misleading. However, a measure of center paired with an appropriate measure of variability and display tells the story of the data.

Education Standard

NCTM Principals and Standards

Select and use appropriate statistical methods to analyze data

- Find, use, and interpret measures of center and spread, including mean and interquartile range.

I can statement

I can describe data using measures of center and spread.

I can display single variable data with a box plot.

Introduction

Dr. Nic's Math and Stats video: *Understanding Summary statistics*

https://creativemaths.net/videos/video-mean/

Notes and Examples

In addition to the shape of a set of numbers, descriptive statistics is concerned with location and spread.

A *Measure of Central Tendency* is a summary statistic that describes the *location of the center* or typical value of a set of numbers.

Mean – is the arithmetic average of a set of numbers. $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$

Median – is the middle number of a set of numbers

Mode – is the response most often observed

The **median** is the middle value in a set of numbers. To find the median, arrange your data from smallest to largest and find the number that splits the data set in half.

In example A, the size of the data set is an *odd* number. The median in example A is 12. There are six values less than 12 and six values greater than 12.

In example B, the size of the data set is an *even* number. The median is 28, the average (mean) of the two middle numbers, 27 and 29.

The median is a *robust* measure of center. This means that it is not influenced by extreme values. In example C, the median is 46. If you replace the largest four values with much larger values, the median is unchanged.

A.

Median Odd
23
21
18
16
15
13
12
10
9
7
6
5
2

B.

	Median Even
	40
	38
	35
	33
	32
	30
28	29
	27
	26
	24
	23
	22
	19
	17

C.

Median	Median Fixed
69	112
56	93
54	89
52	82
47	47
46	46
46	46
45	45
43	43
36	36
35	35
34	34
31	31

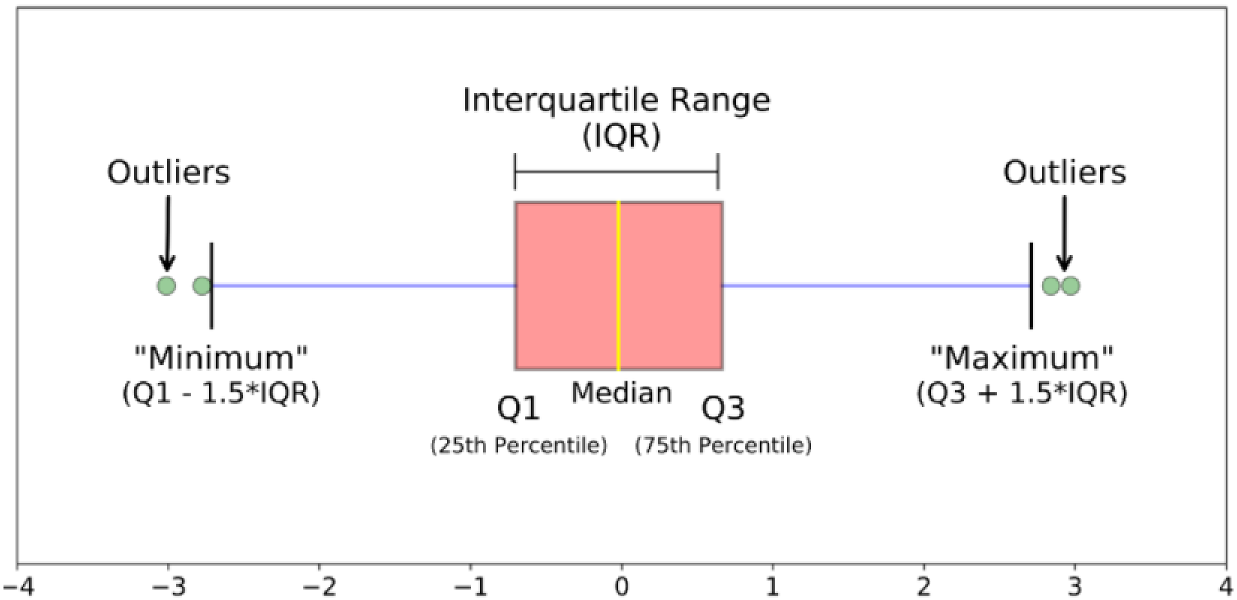
Calculate the mean for the examples above:

A. $\bar{x}$ = __________ B. $\bar{x}$ = _28.214_______ C. $\bar{x}$ = _______

$\bar{x}$ = _______

Notice that in example C, unlike the median, the **mean** value *is* sensitive to extreme values.

A measure of center is more meaningful when it is paired with a measure of spread. A **Box Plot** effectively displays the **median** and the **range** of a set of numbers.



<u>Measures of Location</u>
Minimum
Quartile 1 (Q1) – 25th percentile
Median - 50th percentile
Quartile 3 (Q3) – 75th percentile
Maximum

<u>Measures of Spread</u>
Range: Maximum – Minimum
Interquartile range (IQR): Quartile 3 – Quartile 1

An **Outlier** is an observation that is very different from the rest of the data. An observation may be considered an outlier if it is below the *lower fence* (Q1 – 1.5•IQR) or above the *upper fence* (Q3 + 1.5•IQR).

In the space below, construct a box plot for the adjusted data in example C from above.

Formative assessment

Use the data in example B and subtract ten from the four smallest values. That is, replace 23 with 13, 22 with 12, 19 with 9, and 17 with 7. How will this change to the data effect the mean of the data?

____

Check your answer by calculating the mean of the adjusted data from example B. $\bar{x}$ = _________

Answers to Formative assessment

Use the data in example B and subtract ten from the four smallest values. That is, replace 23 with 13, 22 with 12, 19 with 9, and 17 with 7. How will this change to the data effect the mean of the data?

This will make the mean smaller. It will pull the mean to the left of the median.

Check your answer by calculating the mean of the adjusted data from example B. $\bar{x} = 25.357$

Author

Martha J. Nelson

Acknowledgements

This curriculum was developed under the National Science Foundation RET grant #1953102. However, these contents do not necessarily represent the policies of the National Science Foundation, and you should not assume endorsement by the federal government.

Exercise #2 Analyzing Single Variable Data

The Data:

The data in for this exercise was collected from a model of the intersection of Labeaux and County Road 18 in Albertville MN created using PTV VISSIM software. Videos of traffic simulation which generated the data:

0% AV saturation: https://www.youtube.com/watch?v=IkwWuVd-EAI
90% AV saturation: https://www.youtube.com/watch?v=S3Czr_taME0
3D view: https://www.youtube.com/watch?v=ylzQreGefqY

The Variables:

The software collected data on the following variables:

Average Vehicle Delay is the average number of seconds a vehicle is stopped at the intersection.

Average Queue Length is the average length in meters of the line of vehicles stopped.

Maximum Queue Length is the maximum length in meters of the line of vehicles stopped.

Number of Queue Stops is the number of stops made by all vehicles at the intersection.

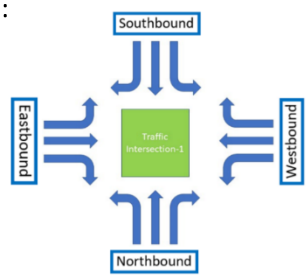
These variables are measured at each entrance to the intersection:

Westbound **(WB)**

Southbound **(SB)**

Eastbound **(EB)**

Northbound **(NB).**



Each variable is measured for seven autonomous vehicle saturation rates:

0% AV – all human driven cars

15% AV – 85% human driven cars

30% AV – 70% human driven cars

45% AV – 55% human driven cars

60% AV – 40% human driven cars

75% AV – 25% human driven cars

90% AV – 10% human driven cars

Each simulation lasted 90 minutes (5400 seconds) with data collected every 15 minutes (900 seconds) and the first and last 15-minute intervals were discarded. Therefore, the data is collected for four time intervals for each run of the simulation:

900-1800 seconds, **1800-2700** seconds, **2700-3600** seconds, **3600-4500** seconds

At each AV saturation rate, the simulation was run ten times. As a result, at each variable at each AV saturation rate there are 160 observations. One for each of the four directions at each of the four time intervals for each of the ten simulations.

At each saturation rate, the autonomous vehicles are tested using three different driving behaviors:

Cautious, **Normal**, **Aggressive**

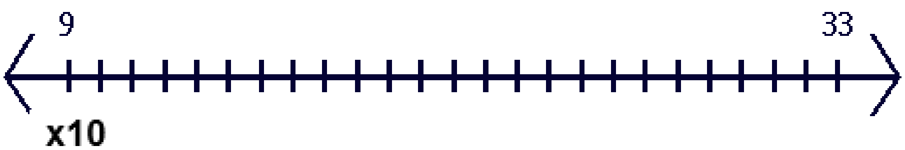
	Queue stops									
	0% AV									
Run	TIMEINT		WB	QSTOPS	SB	QSTOPS	EB	QSTOPS	NB	QSTOPS
1	900-1800		1	123	2	312	3	113	4	274
1	1800-2700		1	104	2	299	3	100	4	279
1	2700-3600		1	127	2	234	3	124	4	297
1	3600-4500		1	157	2	247	3	142	4	228
2	900-1800		1	139	2	258	3	173	4	274
2	1800-2700		1	184	2	258	3	116	4	269
2	2700-3600		1	152	2	285	3	119	4	263
2	3600-4500		1	141	2	303	3	102	4	208
3	900-1800		1	117	2	271	3	134	4	252
3	1800-2700		1	116	2	291	3	117	4	235
3	2700-3600		1	107	2	249	3	134	4	280
3	3600-4500		1	126	2	233	3	110	4	207
4	900-1800		1	105	2	228	3	103	4	274
4	1800-2700		1	135	2	273	3	120	4	271
4	2700-3600		1	110	2	289	3	141	4	297
4	3600-4500		1	131	2	329	3	121	4	257
5	900-1800		1	176	2	267	3	122	4	301
5	1800-2700		1	138	2	248	3	131	4	262
5	2700-3600		1	167	2	280	3	128	4	278
5	3600-4500		1	141	2	312	3	139	4	289
6	900-1800		1	128	2	254	3	117	4	237
6	1800-2700		1	104	2	243	3	102	4	294
6	2700-3600		1	145	2	301	3	95	4	249
6	3600-4500		1	157	2	246	3	124	4	294
7	900-1800		1	107	2	311	3	139	4	280
7	1800-2700		1	148	2	262	3	135	4	320
7	2700-3600		1	113	2	289	3	102	4	253
7	3600-4500		1	132	2	275	3	108	4	267
8	900-1800		1	148	2	238	3	120	4	285
8	1800-2700		1	130	2	252	3	165	4	294
8	2700-3600		1	143	2	306	3	129	4	250
8	3600-4500		1	122	2	285	3	158	4	255
9	900-1800		1	141	2	240	3	150	4	246
9	1800-2700		1	133	2	264	3	144	4	275
9	2700-3600		1	140	2	269	3	145	4	328
9	3600-4500		1	143	2	233	3	120	4	256
10	900-1800		1	125	2	244	3	165	4	257
10	1800-2700		1	135	2	278	3	144	4	258
10	2700-3600		1	132	2	265	3	122	4	289

10	3600-4500		1	134	2	277	3	127	4	239

Using the queue stop data for 0% AV at each intersection entrance direction:

1. Sort the data from smallest to largest.
2. Calculate the Median and the Mean.
3. Construct a Boxplot.
4. Report the Range and the Interquartile Range (IQR).

Westbound:



Mean: ______________________________

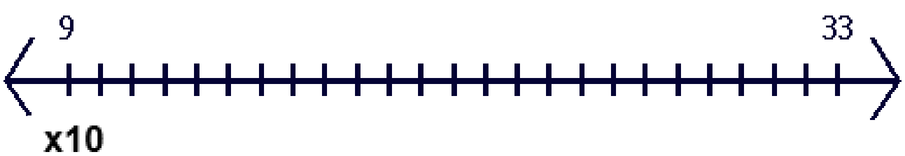
5-number summary: ________________

Range: _____________________________

Median: ____________________________

IQR:________________________________

Southbound:



Mean: ______________________________

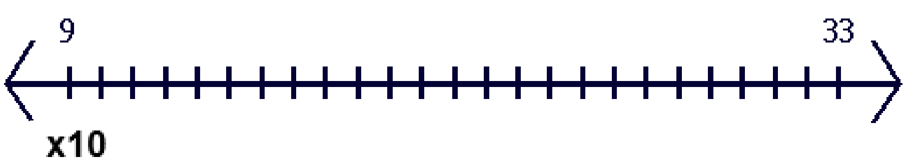
5-number summary: ________________

Range: _____________________________

Median: ____________________________

IQR:________________________________

Eastbound:



Mean: ______________________________

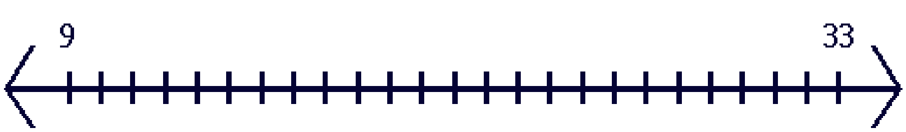
5-number summary: ________________

Range: _____________________________

Median: ____________________________

IQR:________________________________

Northbound:



Mean: ______________________________

5-number summary: ________________

Range: _____________________________

Median: ____________________________

IQR:_______________________________

x10 cise#2 Analyzing Single variable Data Answers

Using the queue stop data for 0% AV at each intersection entrance direction:

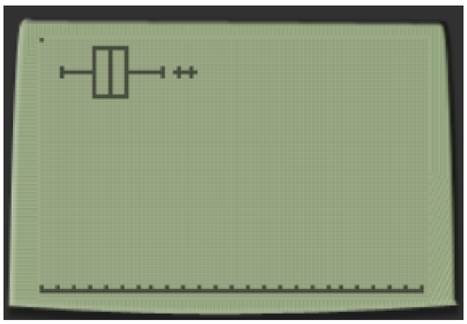
1. Sort the data from smallest to largest.

2. Calculate the Median and the Mean.

3. Construct a Boxplot.

4. Report the Range and the Interquartile Range (IQR).

Westbound:



Mean: 133.9

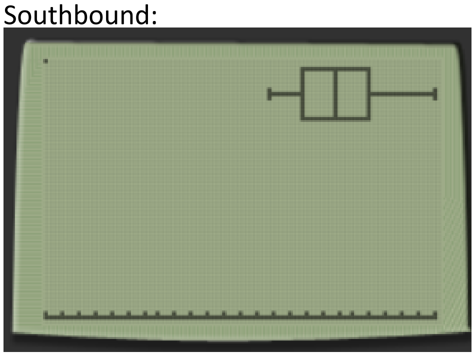
5-number summary: (104, 122.5, 133.5, 143, 184)

Range: 80

Median: 133.5

IQR: 20.5

Southbound:



Mean: 269.95

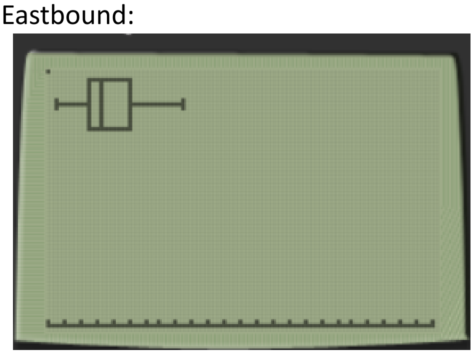
5-number summary: (228, 247.5, 268, 289, 329)

Range: 101

Median: 268

IQR: 41.5

Eastbound:



Mean: 127.5

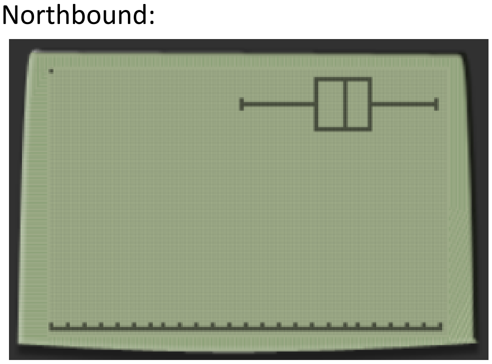
5-number summary: (95, 116.5, 124, 140, 173)

Range: 78

Median: 124

IQR: 23.5

Northbound:



Mean: 268.025

5-number summary: (207, 252.5, 270, 287, 328)

Range: 121

Median:270

IQR: 34