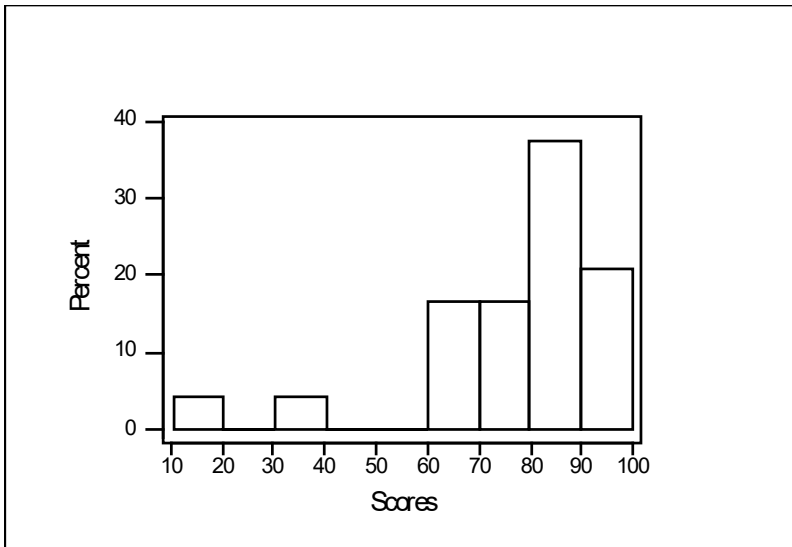


CASE : Statistics Test Scores

- A sample of students obtained the following scores on a statistics tests.

65	13	75	89	92	73	82	85
92	87	95	62	79	82	91	87
65	81	92	86	31	63	74	85



What are the main features of this distribution?

Time Plots

TIME PLOT

A **time plot** of a variable plots each observation against the time at which it was measured. Always put time on the horizontal scale of your plot and the variable you are measuring on the vertical scale. Connecting the data points by lines helps emphasize any change over time.

TREND AND SEASONAL VARIATION

A **trend** in a time series is a persistent, long-term rise or fall.

A pattern in a time series that repeats itself at known regular intervals of time is called **seasonal variation**.

Time Plots

yahoo!finance

Search for news, tickers or companies



News

Finance

Sports

My Portfolio

News

Markets

Research

Personal Finance

Videos

Watch Now

Summary

News

Chart

Community

Historical Data

Options

Components

SNP - Delayed Quote • USD

S&P 500 (^GSPC)

☆ Follow

6,637.97 -18.95 (-0.28%)

At close: 4:45:27 PM EDT

1D 5D 1M 6M YTD 1Y 5Y All

Key Events

Mountain

Advanced Chart



37,487.60%

6,637.97

5,000.00

2,500.00

0.00

1990

1995

2000

2005

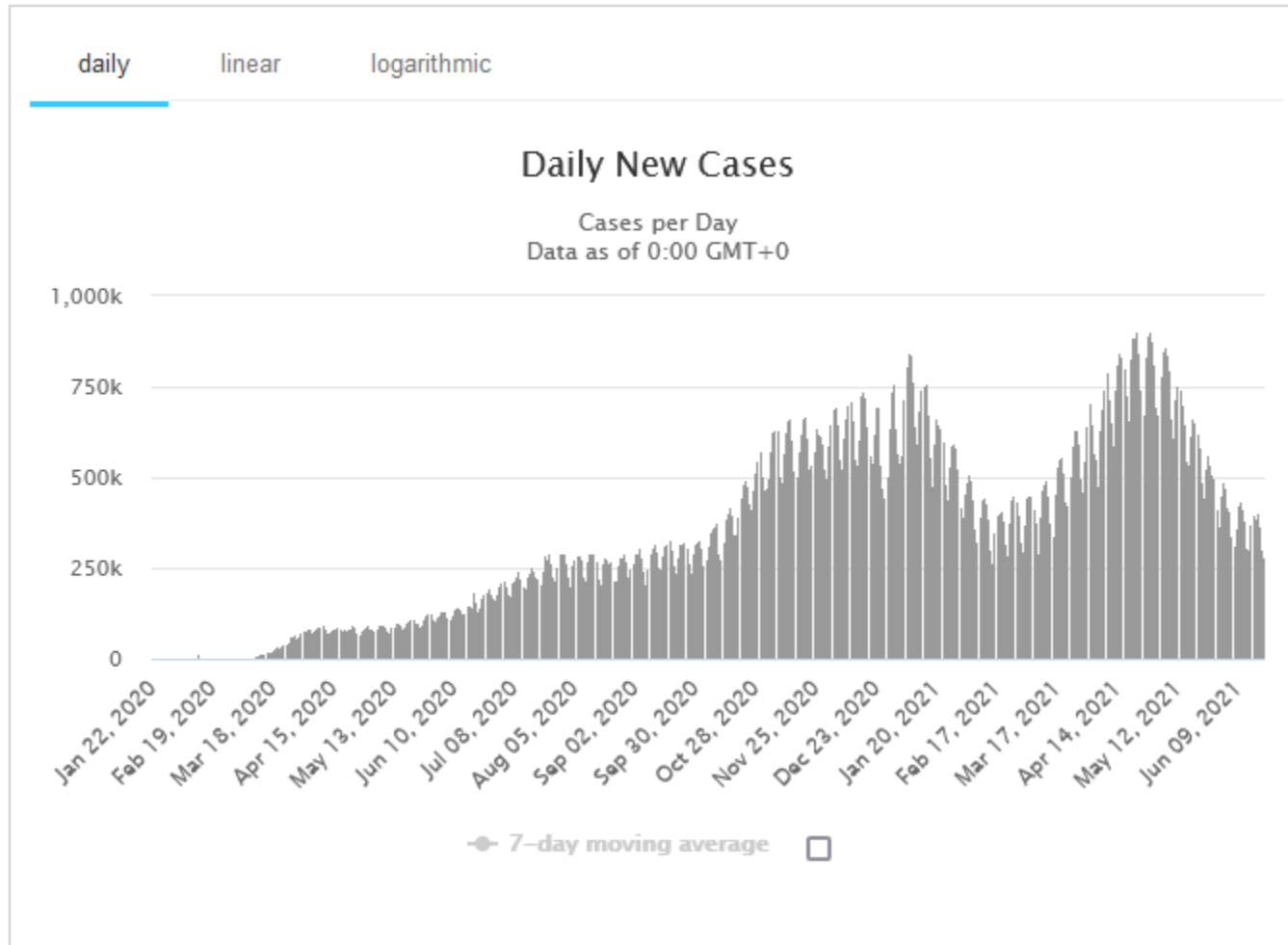
2010

2015

2020

2025

Time Plots



CASE : Statistics Test Scores

- A student received the following scores on the 5 statistics tests during a quarter

45 67 78 81 95 😊

- Another student in same class scored:

95 81 78 67 45 😞

- A third student scored:

67 95 78 45 81 😐

Note: A histogram does not capture the important information in the sequence of the data

Characterizing (Quantitative) Data Using Numerical Methods

Measures of Central Tendency

- Sample Mean
- Sample Median
- Mode

Measures of Variation

- Range
- Sample Variance or Standard Deviation
- Inter Quartile Range (IQR)

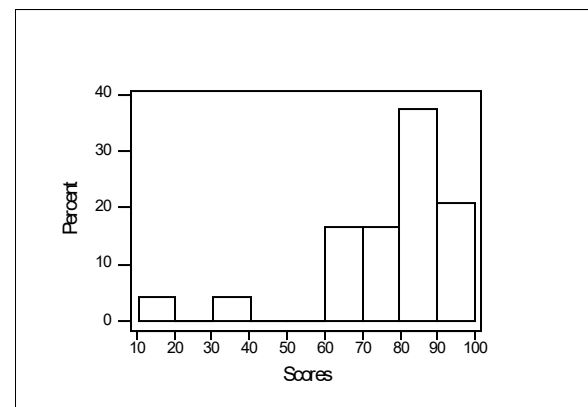
Sample Mean

The Sample Mean (often referred to as average) is the sum of all the observations divided by the number of observations n :

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

Notes:

- The mean is the balance point of the histogram
- The population mean μ is the mean of all x values for the entire population
- We usually cannot measure μ but would like to estimate its value



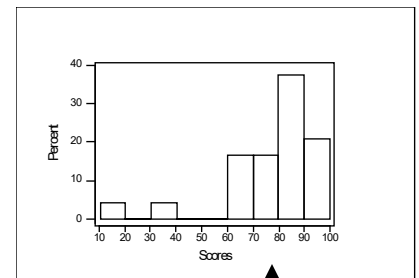
CASE : Statistics Test Scores

- A sample of students obtained the following scores on a statistics tests.

65	13	75	89	92	73	82	85
92	87	95	62	79	82	91	87
65	81	92	86	31	63	74	85

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

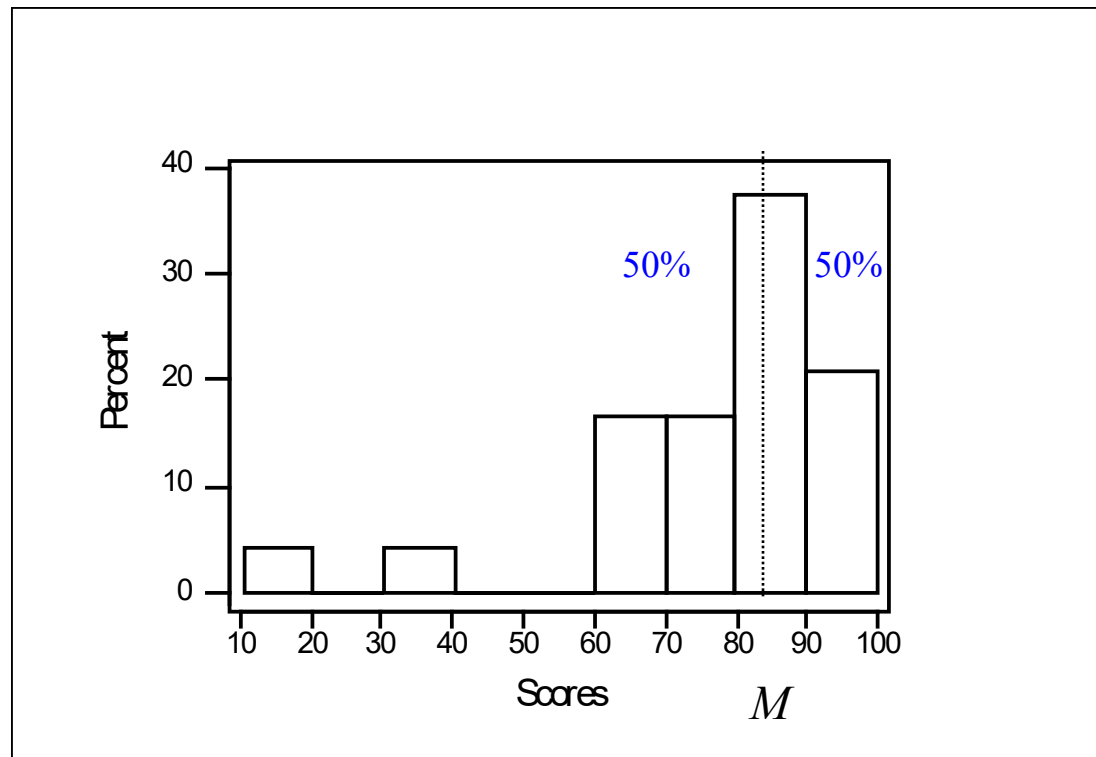
$$= \frac{1}{24} (65 + 92 + 65 + \dots + 85 + 87 + 85) = \frac{1826}{24} = 76.08$$



NOTE: This measure is sensitive to outliers

Sample Median

The [Sample Median](#) M represents the middle observation chosen such that at least 50% of the observations are less than or equal to M and at least 50% are greater than or equal to M .



Sample Median

Procedure for finding the median

1. Arrange observations in order from smallest to largest
2. Identify the location of the median = $(n+1)/2$
3. If n is odd then M is observation # $(n+1)/2$ in the ordered list.
4. If n is even then M is the average of the two observations closest to # $(n+1)/2$

Example:	5	10	19	2	13	
In Order :	2	5	10	13	19	$(n+1)/2=3$

$$M = 10$$

CASE : Statistics Test Scores

- A sample of students obtained the following scores on a statistics test:

65 13 75 89 92 73 82 85
 92 87 95 62 79 82 91 87
 65 81 92 86 31 63 74 85

In order:

13	31	62	63	65	65	73	74
75	79	81	82	82	85	85	86
87	87	89	91	92	92	92	95

$(n+1)/2 = 12.5$

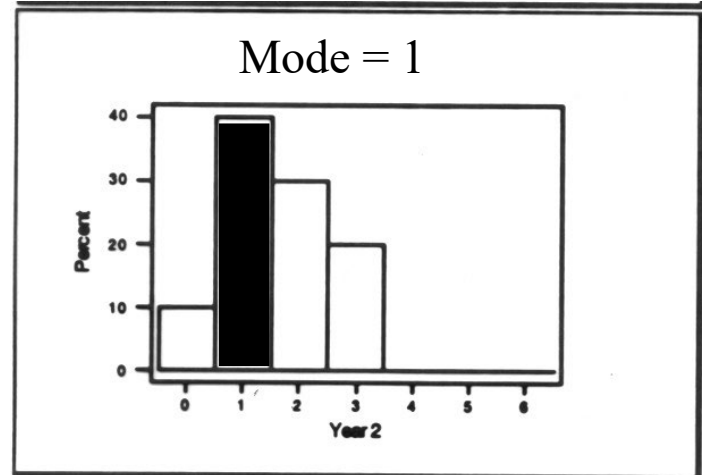
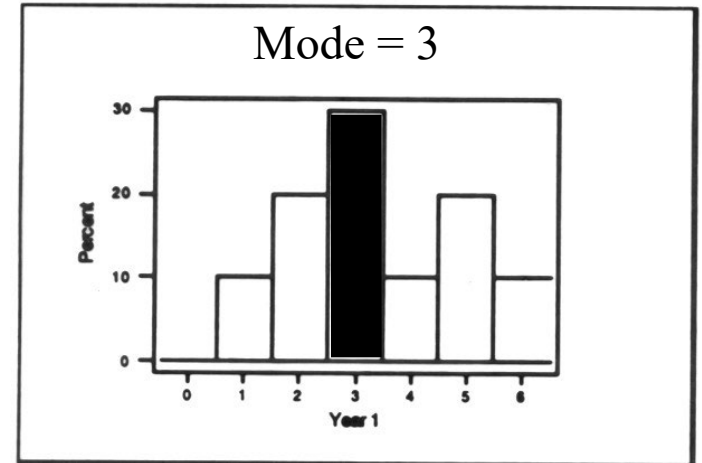
$M = 82$

Stem	Leaf
1	3
2	
3	1
4	
5	
6	2 3 5 5
7	3 4 5 9
8	1 2 2 5 5 6 7 7
9	1 2 2 2 5

Mode

Mode = The most frequent observation

# Service Visits	Frequency		Relative Frequency	
	1st year	2nd year	1st year	2nd year
0	0	1	0	0.1
1	1	4	0.1	0.4
2	2	3	0.2	0.3
3	3	2	0.3	0.2
4	1	0	0.1	0
5	2	0	0.2	0
6	1	0	0.1	0
Total	10	10	1	1



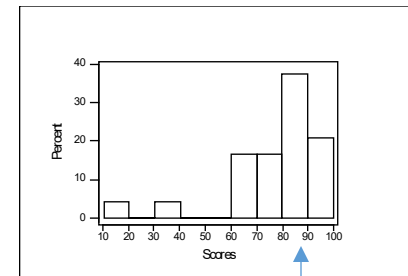
CASE : Statistics Test Scores

Mode?

Mode = ~~92~~ 82

65	13	75	89	92	73	82	85
92	87	95	62	79	82	91	87
65	81	92	86	31	63	74	85

+1 +1



Note: Mode is not a good measure of central tendency when most observations occur only once or twice. A modal class (interval) may be used instead.

Mean or Median?

- At “Jerry’s Burgers” they pay a mean (average) hourly salary of \$25.32/hour. The median salary is \$16.66/hour. Which of the two measures is most informative if you were to apply for a part-time job there?
- Suppose you were moving to a foreign country and you want to know something about typical house prices. Would you rather know the mean of the median house price?
- Suppose it is reported that an American woman spends on average \$10.50 per day at department stores and for men that same figure is \$7.25 per day. The median in each case is \$0.00. Which measure is most useful for comparing men and woman’s shopping habits?



Characterizing (Quantitative) Data Using Numerical Methods

Measures of Central Tendency

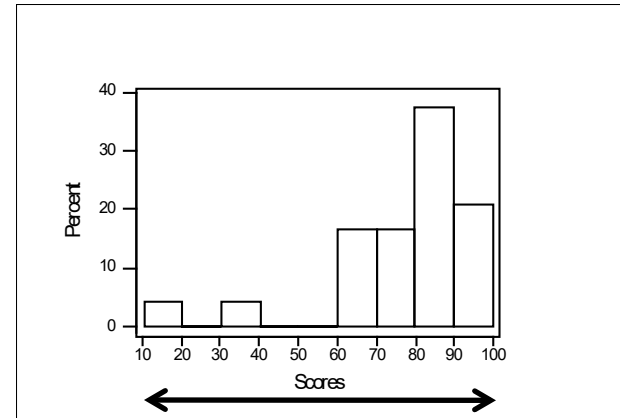
- Sample Mean
- Sample Median
- Mode

Measures of Variation

- Range
- Sample Variance or Standard Deviation
- Inter Quartile Range (IQR)

Range

$$\text{Range} = \text{Max} - \text{Min}$$



65	13	75	89	92	73	82	85
92	87	95	62	79	82	91	87
65	81	92	86	31	63	74	85

$$\text{Range} = 95 - 13 = 82$$