Means and Medians

<u>Mean</u> is an arithmetic average.

Compute by adding all the data values and dividing by the number of data values

<u>Median</u> is the value in the middle of all the data values.

Compute by lining all the data values up in increasing order and finding the one in the middle by counting in from each end

Median is the half-way point

Note: if the number of data values is even, then the median is the average of the two values on either side of the middle

(ex. If there are 12 data values, then the median is the average of the 6^{th} and the 7^{th} data values)

1. Find the mean and median of the following data set:

{3, 9, 16, 1, 25}

mean: (3+9+16+1+25)/5 = 10.8

median: In order, 1, 3, 9, 16, 25

There are 5 data values, so the third is the median = 9

2. Find the mean and median of the following data set:

{35, 74, 11, 48}

mean: (35+74+11+48)/4 = 42

median: In order, 11, 35, 48, 74

There are 4 data values, so the median is the middle:

the average of 35 and 48 = 41.5

3. Find the mean and median of the following data, represented in a frequency table:

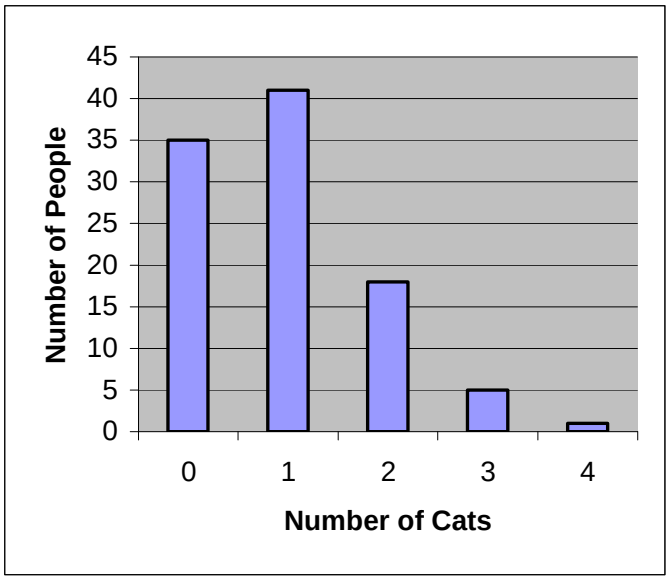
# of cats	0	1	2	3	4
# of people with that many cats	35	41	18	5	1

mean: How many cats total?

0x35 + 1x41 + 2x18 + 3x5 + 4x1 = 96

96/100 = .96

So the average number of cats owned is .96 (slightly less than one cat per person)



median: The data is already in order, but it is grouped. There are 100 data values and we are looking for the middle value. Since 100 is even, we will need to average the 50^{th} and 51^{st} data values. Find the 50^{th} and 51^{st} values by counting from both ends toward the middle.

Counting from the ends puts us in the column associated with 1. Thus the 50^{th} person has 1 cat, and the 51^{st} person has 1 cat. The average of 1 and 1 is 1, so the median is 1.

Survey results for number of rooms in a house

# of rooms	2	4	6	8	10	12
# of houses	2	7	12	18	4	5

mean: (2x2 + 4x7 + 6x12 + 8x18 + 10x4 + 12x5)/48 = 7.25

median: Find the middle value. First, how many data values do we have? 48

Find the middle: 48 is even, counting in from each end we will be needing the 24^{th} and the 25^{th} data values to average.

Counting from the left, 2 + 7 + 12 = 21 and three more will put us in the column with 8 rooms.

Counting from the right, 12 + 10 = 22, and 2 more will put is in the column with 8 rooms also.

The median is the average of 8 and 8, so the median is 8.

If you lined up the data, it would look like:
22444444666666666666888888888888888888 10 10 10 10 12 12 12 12 12
so you could count over 24/25 and you would find the median is 8. This, however, is tedious, and it is easier to use the grouped chart).

Percentiles- give more information about how the data is distributed

The p^{th} percentile- p% of the numbers are at or below this value

Example: SAT score

If score of 605 is 79^{th} percentile, say:

79% of test takers scored 605 or less
21% of test takers scored 605 or more

50^{th} percentile = median

Using mean and median:

Suppose average salary of philosophy graduates is $5,730,000.

Is it as good as it seems to be a philosophy major?

More information: one major is a working as a professional basketball player making $3.5 million.

If we take him out of the picture, the average salary is closer to $30, 135.

Another example:

Find the mean and median of the following set of exam scores:

{24, 77, 78, 83, 87, 88, 89, 92, 93}

Mean: 79 (looks like most students got C/B)
Median: 87 (better reflects the higher grades)

The large salary of $3.5 million and the low test score of 24 are called "outliers" and they affect the average value. The median is not as affected by the outliers.

Know more about the distribution:

Quartiles (note: multiple methods to do this- follow the book)

Median (M) is 50^{th} percentile (or 2^{nd} quartile)
Q_1 is the first quartile (25%)
Q_3 is the third quartile (75%)

Example: {24, 77, 78, 83, 87, 88, 89, 92, 93}

Find Q_1. If you have 9 items, take .25 x 9=2.25, round up to 3. Q_1 is the 3^{rd} data value, or 78.

Find Q_3. If you have 9 items, take .75 x 9=6.25, round up to 7. Q_3 is the 7^{th} data value, or 89.

***Min* = smallest value, or 24**
***Max* = largest value, or 93**

Five-Number Summary:

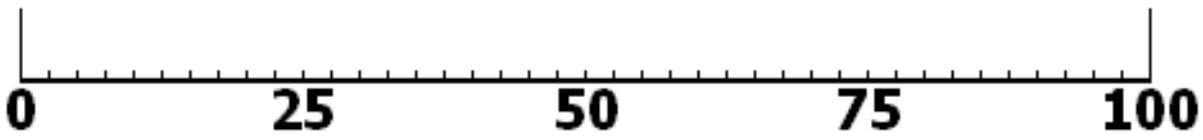
Min **– smallest value**
Q_1 – first quartile
M – median
Q_3 – third quartile
Max **– largest value**

Box Plot- picture of five-number summary

If:

Min **– 24**
Q_1 – 78
M – 87
Q_3 – 89
Max **– 93**

Draw the box plot:



Data Distribution- Spread

Standard Deviation- measures the spread of data, or deviation from mean

The larger the standard deviation, the more spread-out the data

Distribution for Exam 1

Statistics: Exam 1

Graded out of: 0.00 Highest grade: 100.00 Mean grade: 82.68 Standard deviation: 13.39

Number of records: 442 Lowest grade: 25.00 Median grade: 88.00

Score Range	Frequency
[0, 10)	
[10, 20)	
[20, 30)	2
[30, 40)	2
[40, 50)	4
[50, 60)	26
[60, 70)	37
[70, 80)	98
[80, 90)	117
[90, 100)	122
[100]	34

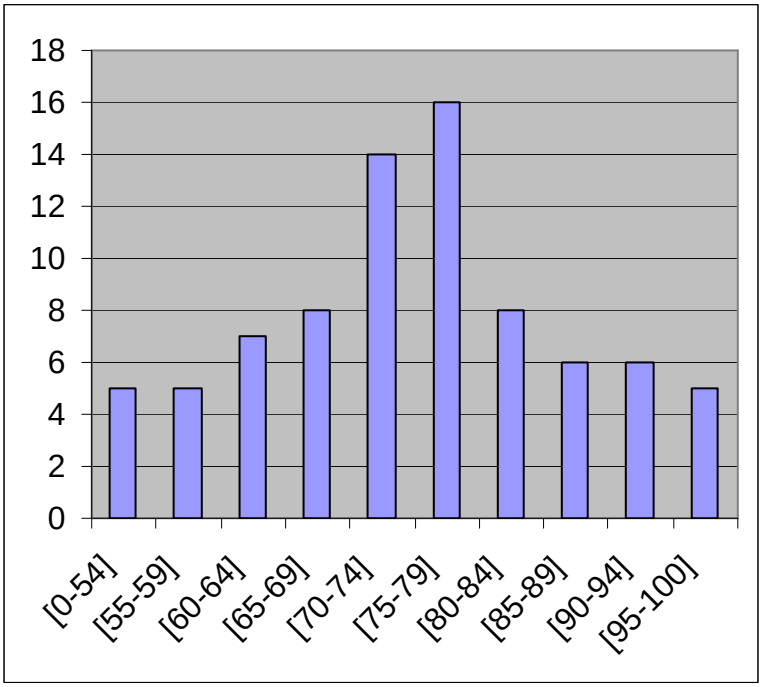
Results from Exam 1.

Mean is 82.68
Median is 88
Standard deviation is 13.39

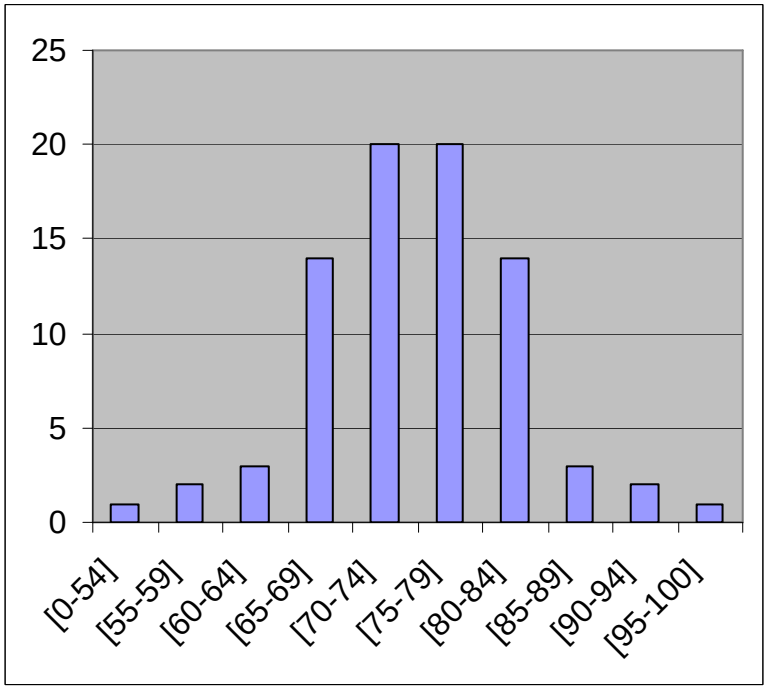
The Standard Deviation describes how spread out the data is.

Consider the two examples with two different standard deviations. The chart on top has a larger standard deviation than the chart on bottom.

Larger Standard Deviation:



Smaller Standard Deviation



Example: Homework Scores

Angela's Homework Scores:

85, 86, 87, 88, 89, 91, 92, 93, 94, 95

Mean = 90 points
Standard Deviation = 3.3 points

Means that most of the homework scores were within 3.3 points of the mean.

Most of the homework was between 86.7 and 93.3

What is the standard distribution for Sally's homework scores?

100, 100, 100, 100, 100, 100, 100, 100, 100, 100

All the data is the same
the standard distribution is 0.

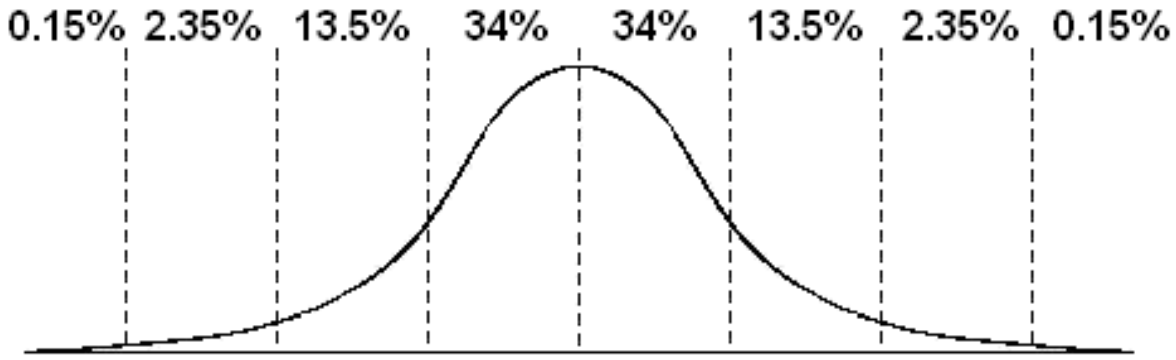
Susan's Homework Scores:

5, 15, 25, 35, 45, 55, 65, 75, 85, 95

Standard Deviation is almost 29 points.

Normal Distribution

Normal Distribution: distribution of data that has a "bell" shape. Data is balanced around the mean.



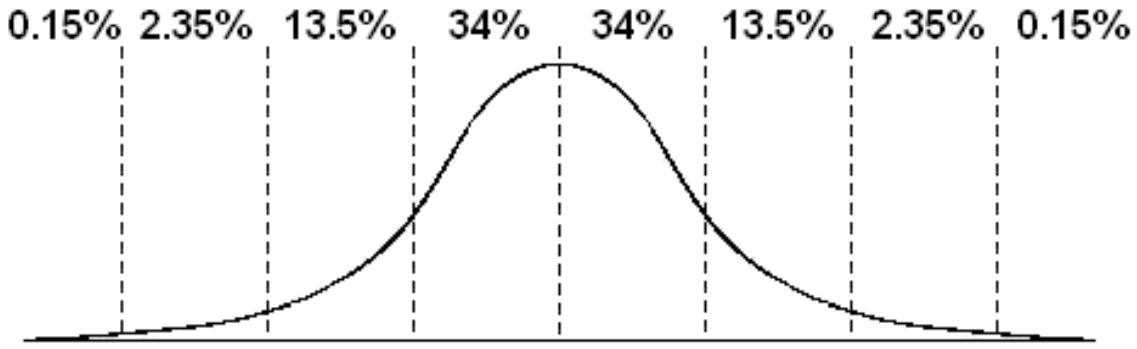
The mean, or average, describes the middle of this data.

In a normal distribution, the mean = median

<u>The Empirical Rule (or the 68-95-99.7 Rule):</u>
68% of the data is always within one standard deviation of the mean. 95% is always within 2 standard deviations, and 99.7% is within 3 standard deviations.

Measure the body length of 100 parasitic wasps The data is normally distributed with mean 1.46 cm and standard deviation 0.2 cm.

Fill in the picture:



How many wasps are shorter than 1.26 cm?

How many wasps are longer than 1.86 cm?

What percentage of wasps are at least 1.66 cm?

What percentage of wasps are shorter than .86 cm or longer than 2.06 cm?