8-1 Measures of Central Tendency and Variation

Extension: Data Distributions

Essential question: *How can you use shape, center, and spread to characterize a data distribution?*

COMMON CORE Standards for Mathematical Content

CC.9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).*

CC.9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).*

Recall that the *mean*, *median*, and *mode* are measures of central tendency—values that describe the center of a data set.

The *mean* is the sum of the values in the set divided by the number of values. It is often represented as $\overline{x}$.

The *median* is the middle value or the mean of the two middle values when the set is ordered numerically.

The mode is the value or values that occur most often. A data set may have one mode, no mode, or several modes.

many

Find the mean, median, and mode of the data set.

{6, 9, 3, 8}

$\boxed{\text{mean}}$

$$= \frac{6+9+3+8}{4} = \frac{26}{4} = 6.5$$

$\boxed{\text{median}}$

3, 6, 8, 9

$$\frac{6+8}{2} = \frac{14}{2} = 7$$

$\boxed{\text{mode}}$

None

For numerical data, the weighted average of all of those outcomes is called the **expected value** for that experiment.

The **probability distribution** for an experiment is the function that pairs each outcome with its probability.

The probability distribution of successful free throws for a practice set is given below. Find the expected number of successes for one set.

Number of Good Free Throws, n	0	1	2	3
Prob. of n Good Free Throws	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{1}{5}$	$\frac{1}{2}$

$$0\left(\frac{3}{20}\right) + 1\left(\frac{3}{20}\right) + 2\left(\frac{1}{5}\right) + 3\left(\frac{1}{2}\right)$$

$$0 + \frac{3}{20} + \frac{2}{5} + \frac{3}{2}$$

$$\frac{3}{20} + \frac{8}{20} + \frac{30}{20} = \frac{41}{20} = 2.05$$

The probability distribution of the number of accidents in a week at an intersection, based on past data, is given below. Find the expected number of accidents for one week.

Number of accidents n	0	1	2	3
Probability of n accidents	0.75	0.15	0.08	0.02

$$0(0.75) + 1(0.15) + 2(0.08) + 3(0.02)$$

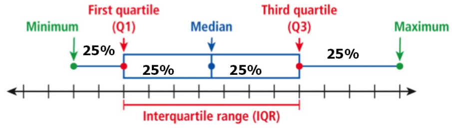
$$0.15 + 0.16 + 0.06$$

$$= 0.37$$

The data sets **{19, 20, 21}** and **{0, 20, 40}** have the same mean and median, but the sets are very different. The way that data are spread out from the mean or median is important in the study of statistics.

A *measure of variation* is a value that describes the spread of a data set. The most commonly used measures of variation are the *range*, the interquartile range, the *variance*, and the *standard deviation*.

A *box-and-whisker plot* shows the spread of a data set. It displays 5 key points: the **minimum** and **maximum** values, the **median**, and the **first** and **third quartiles**.



The quartiles are the medians of the lower and upper halves of the data set. If there are an odd number of data values, do not include the median in either half.

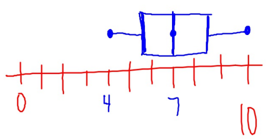
The *interquartile range*, or IQR, is the difference between the 1st and 3 rd quartiles, or Q3 – Q1. It represents the middle 50% of the data.

Make a box-and-whisker plot of the data. Find the interquartile range.

{~~6~~, ~~8~~, ~~7~~, ~~5~~, 10, ~~6~~, 9, ~~8~~, ~~4~~}

4, 5, 6, 6, 7, 8, 8, 9, 10

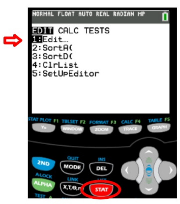
$LQ = \frac{5+6}{2} = 5.5$ median $UQ = \frac{8+9}{2} = 8.5$



IQR = UQ − LQ

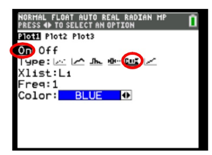
8.5 − 5.5

3



STAT: Edit
enter your data on L1

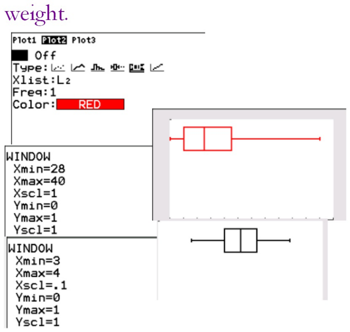
Turn on STAT Plots



Baby	Birth month	Birth weight (kg)	Mother's age
1	5	3.3	28
2	7	3.6	31
3	11	3.5	33
4	2	3.4	35
5	10	3.7	39
6	3	3.4	30
7	1	3.5	29
8	4	3.2	30
9	7	3.6	31
10	6	3.4	32
11	9	3.6	33
12	10	3.5	29
13	11	3.4	31
14	1	3.7	29
15	6	3.5	34
16	5	3.8	30
17	8	3.5	32
18	9	3.6	30
19	12	3.3	29
20	2	3.5	28

L2   L1

Make a box and whisker plot on your calculator and describe the shape of the graph. Describe the shape of Mother's age and of the birth weight.

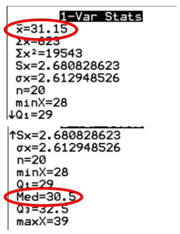


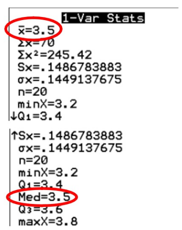
Suppose one of the mothers' ages is chosen at random. Based on the box plot and not the original set of data, what can you say is the approximate probability that the age falls between the median, 30.5, and the third quartile, 32.5? Explain your reasoning.

25% or 0.25 because Q_1, the median, and Q_3 divide the data into four almost-equal parts.

What do you notice about the mean and median for the symmetric distribution (baby weights) as compared with the mean and median for the skewed distribution (mothers' ages)? Explain why this happens.

Mean and median for a symmetric distribution are equal, but mean and median for a skewed distribution are not. This happens because the mean is pulled toward the data values in the longer tail, but the median is not.





For a data set with a first quartile of Q1 and a third quartile of Q3, a value less than Q1 – 1.5(IQR) or greater than Q3 + 1.5(IQR) may be considered to be an outlier. Use this rule to identify any outliers in each data set

Baby	Birth month	Birth weight (kg)	Mother's age
1	5	3.3	28
2	7	3.6	31
3	11	3.5	33
4	2	3.4	35
5	10	3.7	39
6	3	3.4	30
7	1	3.5	29
8	4	3.2	30
9	7	3.6	31
10	6	3.4	32
11	9	3.6	33
12	10	3.5	29
13	11	3.4	31
14	1	3.7	29
15	6	3.5	34
16	5	3.8	30
17	8	3.5	32
18	9	3.6	30
19	12	3.3	29
20	2	3.5	28

Mother's age: 39
Birth weight: none
Birth month: none