1.3 Describing quantative data with numbers

Measures of center

mean, $\bar{x}$, is the sum of the observations divided by the number of observations, n.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

Ex: The following data is the fat content of 8 McDonald's sandwiches:

19, 16, 22, 27, 9, 20, 14, 19

$\bar{x} = (19 + 16 + 22 + 27 + 9 + 20 + 14 + 19) / 8 = 18.25g$

a) Find the mean amount of fat in these sandwiches.

b) the double quarter pounder with cheese has 43 grams of fat. Calculate the mean with this new observations.

$\bar{x} = 21$ g.

Because the mean was significantly influenced by an outlier, it is said not to be a resistant measure. A resistant measure is any measure of center or spread, that is not significantly influenced by outliers.

<u>Median</u>, M, is the midpoint of the distribution.
It is a value such that <u>about</u> ½ of the values are above and below it.

Ex: Find the median of the 1st example.

~~19~~, ~~16~~, ~~22~~, ~~27~~, ~~9~~, ~~20~~, ~~14~~, ~~19~~

9, 14, 16, 19, 19, 20, 22, 27

$\frac{19+19}{2} = 19g$

Ex: Find the Median includ the double quarter pounder w/ cheese.

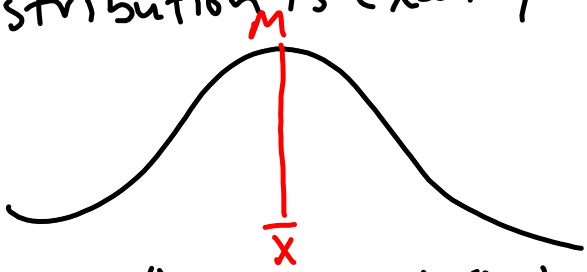
9, 14, 16, 19, 19, 20, 22, 27, 43

$M = 19g$

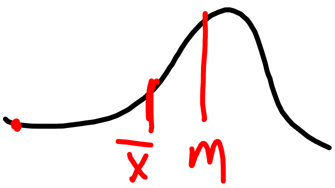
The median is a resistant measure.

Properties of $\bar{x}$ + M

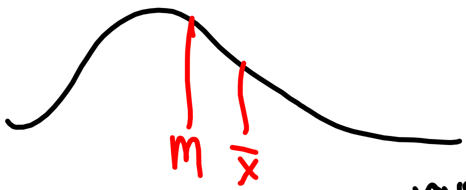
① If a distribution is exactly symmetric.
$\bar{x} = M$



② If the distribution is left skewed, $\bar{x} < M$



③ If the distribution is right skewed, $\bar{x} > M$



④ If the distribution is roughly symmetric, use $\bar{x}$ as the measure of center. If the distribution is strongly skewed, use M.

class 1: 0 , 5 , 10 $\bar{x} = 5$

class 2: 5 , 5 , 5 $\bar{x} = 5$

Measures of Spread

A measure of center is misleading and inappropriate without a corresponding measure of spread. A sound description of a distribution includes a measure of center with its measure of spread.

Range: maximum - minimum value

- is not a resistant measure.

First Quartile, Q_1, is the median of the observations to the left of the median (when its in increasing order)

Third Quartile, Q_3, is the median of the observation to the right of the median.

Interquartile Range (IQR): $Q_3 - Q_1$

- is a resistant measure.

Ex: Calculate the IQR for the McDonalds sandwiches.

$\boxed{9, 14, 16, 19}, 19, \boxed{20, 22, 27, 43}$

$Q_1 = \frac{14+16}{2} = 15$

$Q_3 = \frac{22+27}{2} = 24.5$

$IQR = 24.5 - 15 = 9.5$

Identifying outliers using the 1.5 x IQR rule

an observation that falls more than 1.5 x IQR above the Q_3 or below the Q_1 is a suspected outlier.

Ex: Is the double quarter pounder w/ cheese an outlier?

1.5 x IQR

$1.5 \times 9.5 = 14.25$

$Q_1 - 14.25$

$15 - 14.25 = .75$

$Q_3 + 14.25$

$24.5 + 14.25 = 38.75$

Yes since it has fat content above 38.75 g.

5 number summary

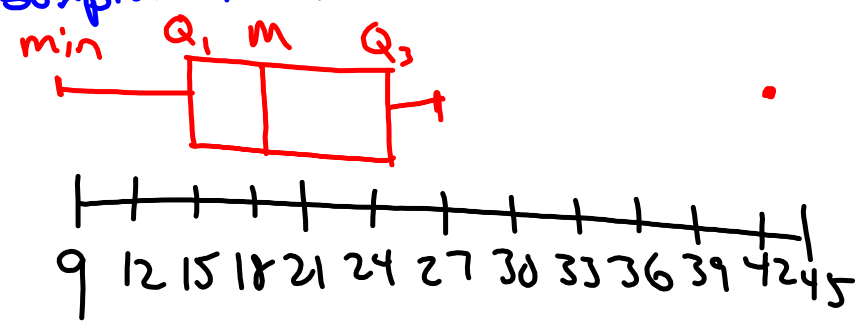
The 5 number summary of a distribution consists of:

1. minimum
2. Q_1
3. Median
4. Q_3
5. maximum

A boxplot is a graphical display of a 5 number summary.

Ex: Make a boxplot for the McDonald's sandwiches.

minimum: 9
$Q_1 = 15$
Median 19
$Q_3 = 24.5$
max = 43



Standard Deviation, s or s_x, measures the typical distance of the values in a distribution from the mean. It is calculated by finding the average deviations (called variance, s_x^2) and find the square root.

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Ex: Find the standard deviation of the McDonald sandwiches.

$$s_x = \sqrt{\frac{(9-21)^2+(14-21)^2+(16-21)^2+(19-21)^2+(19-21)^2+(20-21)^2+(22-21)^2+(27-21)^2+(43-21)^2}{9-1}}$$

$s_x = 9.67$

The typical fat content varies 9.67 grams from the mean of 21 g. in a McDonalds sandwich.

Properties of Standard Deviation

1. Since it measures spread about the mean, it should only be used with the mean!

2. $s_x \geq 0$

3. Has the same units as the original data values.

4. Since the calculation uses the mean, it is NOT a resistant measure.