Measures of Center and Spread

Measures of Center

- Most commonly used measures of center
  - Mode
  - Mean
  - Median
  - Midrange

Mode

- **The value in a distribution for a variable which has the greatest frequency**

Mode

- The value in a distribution for a variable which has the greatest frequency
- *There may be*
  - *No mode,*
  - *One mode, or*
  - *Multiple modes*

Mean

- The mean is also known as
  - the *average*
  - the *arithmetic mean*

Mean

- The mean is also known as
  - the *average*
  - the *arithmetic mean*

- How do we calculate the mean?

Mean

- The mean is also known as
  - the *average*
  - the *arithmetic mean*

- To calculate the mean, we take the *sum of all the data values* and *divide* this sum *by* the *number of data values.*

Sample Mean

- Suppose that we have n data values. Let us represent these data values with a subscripted x to distinguish between the data values, That is, suppose x_1, x_2, x_3, …, x_n are the data.

Then, the sample mean, denoted by $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

Sample Mean

- Suppose x_1, x_2, x_3, …, and x_n are n data values.

  The sample mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Note: n is the sample size.

Sample Mean

- Suppose x_1, x_2, x_3, …, and x_n are n data values.

  Using summation notation, the sample mean can be expressed as

$$\bar{x} = \frac{\sum x_k}{n}$$

where $\sum x_k$ denotes the sum of data values x_1, x_2, x_3, …, and x_n.

Population Mean

- Suppose that we have n data values. Let us represent these data values with a subscripted x to distinguish between the data values, That is, suppose x_1, x_2, x_3, …, x_N are the data.

Then, the population mean, denoted by μ, is given by

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

Population Mean

- Suppose x_1, x_2, x_3, …, and x_N are N data values.

  The population mean, μ, is given by

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

Note: N is the population size.

Population Mean

- Suppose x_1, x_2, x_3, …, and x_N are N data values.

  Using summation notation, the sample mean can be expressed as

$$\mu = \frac{\sum x_k}{N}$$

where $\sum x_k$ denotes the sum of data values x_1, x_2, x_3, …, and x_N.

Mean

- The mean is the *balance point* (center of gravity) for a distribution.
- Visual estimate: place your finger below the point on the horizontal axis of a dot plot or histogram so that you can *balance* it, half the weight to the left and half the weight to the right

Mean

- The mean is the *balance point* (center of gravity) for a distribution.
- Visual estimate: for a distribution that is *approximately normal*, the mean will be directly below the highest point on the bell curve, the tallest stack of dots for a dot plot, and the tallest bar in a histogram

Median

- **The median is the physical middle value for the distribution *when the data values are in numerical order.***

Median

- The median is the physical middle value for the distribution *when the data values are in numerical order*.
- The median divides the distribution into two halves.

Median

- The median is the physical middle value for the distribution *when the data values are in numerical order.*
- The median divides the distribution into two halves.
  - Lower half of the distribution
  - Upper half of the distribution

Median

- The median is the physical middle value for the distribution *when the data values are in numerical order.*
- The median separates the
  - Lower 50% of the distribution

from the

  - Upper 50% of the distribution.

Determining the Median

- Arrange the data in numerical order
- Determine the number of data values
- Count off data values from one end to the middle value

Determining the Median

- If there are an odd number of data values, the median is the middle data value
- If there are an even number of data values, the median is the average of the middle two data values.

Determining the Median

- If there are an odd number of data values, the median is the ***middle data value***

- If there are an even number of data values, the median is the ***average of the middle two data values***.

When should the Mean be Used?

When should the Mean be Used?

- The mean should be used when
  - The distribution is approximately normal (i.e. bell-shaped)
  - The distribution is symmetric around a central value

When should the Mean be Used?

- The mean should be used for a
  - Normal (i.e. bell-shaped) distribution
  - Rectangular (i.e. uniform) distribution
  - A distribution which is symmetric around a central value

When should the Median be Used?

When should the Median be Used?

- The median should be used when a distribution is
  - Skewed-left
  - Skewed-right
  - Not symmetric around a central value

Midrange

- **The midrange is the average of the *smallest* and the *largest* data values.**

Midrange

- **The midrange is the average of the *minimum* and the *maximum* of the distribution for a variable.**

Midrange

- The midrange is the average of the *minimum* and the *maximum* of the distribution for a variable.

- The midrange is a measure of center which is determined by the minimum and maximum values in a distribution for a variable.

Midrange

- The midrange is the average of the *minimum* and the *maximum* of the distribution for a variable.
- The midrange does not take into account any values in the distribution for a variable other than the minimum and maximum.

Range

- The range is the difference between the *largest* and the *smallest* data values.

Range

- The range is the difference between the *maximum* and the *minimum* values of the distribution for a variable.

Range

- The range is the difference between the *maximum* and the *minimum* values of the distribution for a variable.
- The range tells us *nothing* about the center of a distribution for a variable.

Range

- The range is the difference between the *maximum* and the *minimum* values of the distribution for a variable.

- The range tells us *nothing* about the spread of a distribution for a variable about its mean.

Range

- The range is the difference between the *maximum* and the *minimum* values of the distribution for a variable.
- The range tells us *nothing* about the spread of a distribution for a variable about its median.

Range

- The range is the difference between the *maximum* and the *minimum* values of the distribution for a variable.

- The range tells us about the spread of the distribution for a variable.

Measuring Spread Around the Median

- When determining the median, we divide the distribution into halves
  - Dividing each of these halves into halves, we determine the first or lower quartile, Q_1, and the third or upper quartile, Q_3, for the distribution.

Measuring Spread Around the Median

- The first or lower quartile, Q_1, is the median of the lower half of the distribution.
- The third or upper quartile, Q_3, is the median for the upper half of the distribution.

Measuring Spread Around the Median

- The median, Q_1, the first or lower quartile, and Q_3, the third or upper quartile, divide the distribution into quarters.
- That is, the median, Q_1, and Q_3 divide the distribution into four pieces (fourths).

Measuring Spread Around the Median

- The median, Q_1, the first or lower quartile, and Q_3, the third or upper quartile, divide the distribution into quarters.
- That is, the median, Q_1, and Q_3 divide the distribution into four pieces (fourths).
- Note: The median is also known as the second quartile and denoted Q_2.

Measuring Spread Around the Median

- The interquartile range, denoted IQR, is a measure of spread from the lower quartile to the upper quartile,

$$IQR = Q_3 - Q_1$$

Measuring Spread Around the Median

- The interquartile range, denoted IQR, is a measure of spread from the lower quartile to the upper quartile,

$$IQR = Q_3 - Q_1$$

- The IQR is the spread of the middle 50% of the data.

Exploring Spread Around The Mean

- Consider the following numbers:
  3, 5, 1, 8, 0, 7, 3, 6, 5, 2
- Determine the mean.
- Sketch a dot plot.
- Explore the spread around the mean.

Exploring Spread Around The Mean

x_k
3
5
1
8
0
7
3
6
5
2

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	
5	
1	
8	
0	
7	
3	
6	
5	
2	

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	-1
5	1
1	-3
8	4
0	-4
7	3
3	-1
6	2
5	1
2	-2

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	-1
5	1
1	-3
8	4
0	-4
7	3
3	-1
6	2
5	1
2	-2

deviations
or
deviations from the mean

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	-1
5	1
1	-3
8	4
0	-4
7	3
3	-1
6	2
5	1
2	+ -2

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	-1
5	1
1	-3
8	4
0	-4
7	3
3	-1
6	2
5	1
2	+ -2
	0

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$
3	-1
5	1
1	-3
8	4
0	-4
7	3
3	-1
6	2
5	1
2	+ -2
	0

The sum of the deviations from the mean is zero.

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
3	-1	1
5	1	1
1	-3	9
8	4	16
0	-4	16
7	3	9
3	-1	1
6	2	4
5	1	1
2	-2	4

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
3	-1	1
5	1	1
1	-3	9
8	4	16
0	-4	16
7	3	9
3	-1	1
6	2	4
5	1	1
2	-2	4

The squared deviations

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
3	-1	1
5	1	1
1	-3	9
8	4	16
0	-4	16
7	3	9
3	-1	1
6	2	4
5	1	1
2	-2	+ 4

Exploring Spread Around The Mean

x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
3	-1	1
5	1	1
1	-3	9
8	4	16
0	-4	16
7	3	9
3	-1	1
6	2	4
5	1	1
2	-2	+ 4
		62

Exploring Spread Around The Mean

x_k	$x_k-\bar{x}$	$(x_k-\bar{x})^2$
3	-1	1
5	1	1
1	-3	9
8	4	16
0	-4	16
7	3	9
3	-1	1
6	2	4
5	1	1
2	-2	+ 4
		62

The sum of the squared deviations is not zero.

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{\sum \left(x_k - \bar{x}\right)^2}{n - 1}}$$

- Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum \left(x_k - \mu\right)^2}{N}}$$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{\sum \left(x_k - \bar{x}\right)^2}{n - 1}}$$

Used for statistical inference

- Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum \left(x_k - \mu\right)^2}{N}}$$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{n\sum x_k^2 - \left(\sum x_k\right)^2}{n\left(n - 1\right)}}$$

- Population Standard Deviation

$$\sigma = \sqrt{\frac{N\sum x_k^2 - \left(\sum x_k\right)^2}{N^2}}$$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{n\sum x_k^2 - \left(\sum x_k\right)^2}{n\left(n-1\right)}}$$

Used for statistical inference

- Population Standard Deviation

$$\sigma = \sqrt{\frac{N\sum x_k^2 - \left(\sum x_k\right)^2}{N^2}}$$

Standard Deviation

- Use σ, the *population standard deviation*, when you *know* all the values in a population

- Use s, *the sample standard deviation*, when you *have* a random sample *chosen* from the population

Standard Deviation

- Use σ, the *population standard* deviation, when you ***know all the values in a population***
- Use s, *the sample standard* deviation, when you *have* a random ***sample chosen from the population***

Standard Deviation

- Used to measure the spread of the data from the mean
- A measure of the dispersion of a distribution

Standard Deviation

- Used to measure the spread of the data from the mean

- A measure of the dispersion of a distribution

What is "dispersion"?

Standard Deviation

- Used to measure the spread of the data from the mean
- A measure of the dispersion of a distribution

Dispersion is the degree of scatter of data around the mean. …

Standard Deviation

- Used to measure the spread of the data from the mean
- A measure of the dispersion of a distribution

Dispersion is the scattering of the values of a frequency distribution from the mean.

…

Standard Deviation

- Used to measure the spread of the data from the mean
- A measure of the dispersion of a distribution

Dispersion is the spread of a distribution around the central value.

Standard Deviation

- Used to measure the spread of the data from the mean
- A measure of the dispersion of a distribution

Other measures of dispersion include the *semi-interquartile range* and the mean absolute deviation.

Variance

- The variance of a set of values is a measure of variation equal to the square of the standard deviation.

- Variation is a general description of the amount that values vary among themselves. (The terms *dispersion* and *spread* are often used instead of *variation*.)

Which summary statistics should I use to describe a distribution?

Which summary statistics should I use to describe a distribution?

- Use mean and standard deviation when a distribution is
  - Normal (i.e. bell-shaped)
  - Rectangular (i.e. uniform)
  - Symmetric around a central value

Which summary statistics should I use to describe a distribution?

- Use median and quartiles when a distribution is
  - Skewed-left
  - Skewed-right
  - Not symmetric around a central value

So, what should I do first?

- Since we need to know the shape of the distribution in order to determine the distribution type, we should always start by graphing the distribution.

What graph(s) should I use?

What graph(s) should I use?

- What graphs display the shape of the distribution?

What graph(s) should I use?

- What graphs display the shape of the distribution?
  - Dot plots
  - Histograms

Uses for mean and standard deviation for distributions that are not normal

- Suppose you have a representative sample for prices of cars of a particular make or class and you want to use the sample to represent the price of all the cars of this make or class
  - Use the mean

Uses for mean and standard deviation for distributions that are not normal

- Suppose you have a representative sample for prices of cars of a particular make or class and you want to use the sample to represent the price of all the cars of this make or class
  - Use the mean Why?

Uses for mean and standard deviation for distributions that are not normal

- Suppose you have a representative sample for prices of cars of a particular make or class and you want to use the sample to represent the price of all the cars of this make or class
  - Use the mean **Why?**
  - Sample means are approximately normal

Recentering and Rescaling Data

- Recentering – adding the same number to all data values
  - Does not change the shape
  - Does not change the spread
  - "slides" (along horizontal axis for graph) distribution by amount c
    - Changes mean and median by amount c

Recentering and Rescaling Data

- Rescaling – multiplying all data values by same *nonzero* number
  - Does not affect the *basic* shape
  - Stretches or shrinks the distribution
  - IQR multiplied by |d|
  - Mean and median multiplied by d

Influence of Outliers

A summary statistic is

- *Resistant to outliers* if the summary statistic does not change very much if an outlier is removed from a data set

Influence of Outliers

A summary statistic is

- *Sensitive to outliers* if the summary statistic changes when an outlier is removed from a data set

Influence of Outliers

A summary statistic is

- *Sensitive to outliers* if the summary statistic changes when an outlier is removed from a data set
  - Mean, standard deviation, minimum, maximum, midrange, and range are sensitive to outliers

Influence of Outliers

A summary statistic is

- *Sensitive to outliers* if the summary statistic changes when an outlier is removed from a data set
  - Mode, median Q_2, quartiles Q_1 and Q_3, and IQR are less sensitive to outliers

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Value	Frequency
1	27
2	31
3	42
4	40
5	28
6	32

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Value	Frequency
1	27
2	31
3	42
4	40
5	28
6	32

This type of data is also known as *weighted data* as the frequencies are the *weights* for the values.

Sample Mean

- Suppose x_1, x_2, x_3, …, and x_m are data values with corresponding frequencies f_1, f_2, f_3, …, and f_m, respectively.

  The sample mean can be expressed as

$$\bar{x} = \frac{\sum x_k f_k}{\sum f_k}.$$

Sample Mean

- Suppose x_1, x_2, x_3, …, and x_m are data values with corresponding frequencies f_1, f_2, f_3, …, and f_m, respectively.

  The sample mean can be expressed as

$$\bar{x} = \frac{\sum x_k f_k}{n}$$

  for which $n = \sum f_k$.

Population Mean

- Suppose x_1, x_2, x_3, …, and x_m are data values with corresponding frequencies f_1, f_2, f_3, …, and f_m, respectively.

  The sample mean can be expressed as

$$\mu = \frac{\sum x_k f_k}{\sum f_k}.$$

Population Mean

- Suppose x_1, x_2, x_3, …, and x_m are data values with corresponding frequencies f_1, f_2, f_3, …, and f_m, respectively.

  The population mean can be expressed as

$$\mu = \frac{\sum x_k f_k}{N}$$

for which $N = \sum f_k$.

Mean and Standard Deviation from a Frequency Table

x_k	f_k
1	27
2	31
3	42
4	40
5	28
6	32

The outcomes are the values of x_k and the frequencies are the values of $f_{k.}$

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k\ f_k$
1	27	
2	31	
3	42	
4	40	
5	28	
6	32	

A column in which to determine the product of these values, $x_k\ f_k$, is added to the table ...

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k f_k$
1	27	27
2	31	62
3	42	126
4	40	160
5	28	140
6	32	192

A column in which to determine the product of these values, x_k f_k, is added to the table and the products of the values are determined

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k\ f_k$
1	27	27
2	31	62
3	42	126
4	40	160
5	28	140
6	32	192

The mean is the quotient of the sum of the products $x_k\ f_k$ and the sum of the frequencies f_k.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k\ f_k$
1	27	27
2	31	62
3	42	126
4	40	160
5	28	140
6	32	192
	200	707

The mean is the quotient of the sum of the products $x_k\ f_k$ and the sum of the frequencies f_k.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k\ f_k$
1	27	27
2	31	62
3	42	126
4	40	160
5	28	140
6	32	192
	200	707

$$\bar{x} = \frac{707}{200}$$

$$= 3.535$$

$$\approx 3.5$$

The sample mean is the quotient of the sum of the products $x_k\ f_k$ and the sum of the frequencies f_k.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$x_k f_k$
1	27	27
2	31	62
3	42	126
4	40	160
5	28	140
6	32	192
	200	707

$$\mu = \frac{707}{200}$$

$$= 3.535$$

$$\approx 3.5$$

The population mean is the quotient of the sum of the products x_k f_k and the sum of the frequencies f_k.

Mean and Standard Deviation from a Frequency Table

- Suppose each data value x_k occurs with frequency f_k.
- The sample standard deviation of a frequency table is given by

$$s = \sqrt{\frac{\sum \left(x_k - \bar{x}\right)^2 f_k}{\sum f_k - 1}}$$

Mean and Standard Deviation from a Frequency Table

- Suppose each data value x_k occurs with frequency f_k.
- The population standard deviation of a frequency table is given by

$$\sigma = \sqrt{\frac{\sum (x_k - \mu)^2 f_k}{\sum f_k}}$$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{\sum (x_k - \bar{x})^2 f_k}{\sum f_k - 1}}$$

- Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum (x_k - \mu)^2 f_k}{\sum f_k}}$$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{\sum \left(x_k - \bar{x}\right)^2 f_k}{n - 1}} \qquad \text{for } n = \sum f_k$$

- Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum \left(x_k - \mu\right)^2 f_k}{N}} \qquad \text{for } N = \sum f_k$$

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \overline{x})$	$(x_k - \overline{x})^2$	$(x_k-\overline{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

We determine the deviations from the mean.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \bar{x})$	$(x_k - \bar{x})^2$	$(x_k-\bar{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

Then, we square the deviations from the mean to determine the squared deviations from the mean.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \bar{x})$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

We take the product of the frequency and the corresponding value of the squared deviation from the mean.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \bar{x})$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

The quotient of the sum of these values and the sum of the frequencies is used to determine the sample standard deviation and the population standard deviation.

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \bar{x})$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

The square root of the quotient of the sum of these values and *one less than* the sum of the frequencies is the sample standard deviation.

$$s = \sqrt{\frac{521.755}{199}}$$

$$\approx 1.619223401$$

$$\approx 1.6$$

Mean and Standard Deviation from a Frequency Table

x_k	f_k	$(x_k - \bar{x})$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 f_k$
1	27	-2.535	6.426225	173.508075
2	31	-1.535	2.356225	73.042975
3	42	-0.535	0.286225	12.02145
4	40	0.465	0.216226	8.649
5	28	1.465	2.146225	60.0943
6	32	2.465	6.076225	194.4392
	200			521.755

The square root of the quotient of the sum of these values and the sum of the frequencies is the population standard deviation.

$$\sigma = \sqrt{\frac{521.755}{200}}$$

$$\approx 1.61517027$$

$$\approx 1.6$$

Although you may prefer to use the formulas involving the

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{n\sum x_k^2 f_k - \left(\sum x_k f_k\right)^2}{n\left(n-1\right)}}$$

squared deviations from the mean, these formulas are easier to use.

- Population Standard Deviation

$$\sigma = \sqrt{\frac{N\sum x_k^2 f_k - \left(\sum x_k f_k\right)^2}{N^2}}$$

for $n = \sum f_k$ and $N = \sum f_k$

Standard Deviation

- Sample Standard Deviation

$$s = \sqrt{\frac{n\sum x_k^2 f_k - \left(\sum x_k f_k\right)^2}{n\left(n-1\right)}}$$

- Population Standard Deviation

$$\sigma = \sqrt{\frac{N\sum x_k^2 f_k - \left(\sum x_k f_k\right)^2}{N^2}}$$

for $n = \sum f_k$ and $N = \sum f_k$

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Speed	Frequency
42-45	25
46-49	14
50-53	7
54-57	3
58-61	1

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Speed	Frequency
42-45	25
46-49	14
50-53	7
54-57	3
58-61	1

This is interval data. Rather than individual values, the speeds are intervals.

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Speed	Frequency
42-45	25
46-49	14
50-53	7
54-57	3
58-61	1

In order to analyze this data, we need individual values which represent the intervals.

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Speed	Frequency
42-45	25
46-49	14
50-53	7
54-57	3
58-61	1

We can use the midpoint of each interval to represent the interval.

Mean and Standard Deviation from a Frequency Table

- Consider the following data:

Speed	Frequency
42-45	25
46-49	14
50-53	7
54-57	3
58-61	1

The midpoint of an interval is determined by taking the average of the endpoints of the interval.

Mean and Standard Deviation from a Frequency Table

Interval	Midpoint	Frequency
42-45	43.5	25
46-49	47.5	14
50-53	51.5	7
54-57	55.5	3
58-61	59.5	1

The midpoint of an interval is determined by taking the average of the endpoints of the interval.

Mean and Standard Deviation from a Frequency Table

Interval	Midpoint	Frequency
42-45	43.5	25
46-49	47.5	14
50-53	51.5	7
54-57	55.5	3
58-61	59.5	1

Having determined the midpoint for each interval, you can now analyze the data as a frequency table ...

Mean and Standard Deviation from a Frequency Table

Interval	x_k	f_k
42-45	43.5	25
46-49	47.5	14
50-53	51.5	7
54-57	55.5	3
58-61	59.5	1

Having determined the midpoint for each interval, you can now analyze the data as a frequency table for which the midpoints are x_k and the frequencies are f_k.