Measures of Variation

Section 3-2

Jacob Ayers

MAT 110

Lesson #6

Objectives

- Find the range of a data set
- Find the population variance and standard deviation
- Find the sample variance and standard deviation
- Use Chebyshev's Theorem to describe data given a mean and standard deviation
- Use the Empirical Rule to describe data given a mean and standard deviation

Measures of Variation

In order to describe data accurately, we need to know more than the measures of central tendency.

Another important aspect of a data set is its *variation*, or spread.

Measures of Variation

In order to describe data accurately, we need to know more than the measures of central tendency.

Another important aspect of a data set is its *variation*, or spread.

There are three commonly used measures of variation:

- Range
- Variance
- Standard Variation

Measures of Variation

In order to describe data accurately, we need to know more than the measures of central tendency.

Another important aspect of a data set is its *variation*, or spread.

There are three commonly used measures of variation:

- Range
- Variance
- Standard Variation

We will look at each of these measures in turn; again, graphing calculators are very useful when computing them.

Measures of Variation

In order to describe data accurately, we need to know more than the measures of central tendency.

Another important aspect of a data set is its *variation*, or spread.

There are three commonly used measures of variation:

- Range
- Variance
- Standard Variation

We will look at each of these measures in turn; again, graphing calculators are very useful when computing them.

Before we start, let's look at an example to see why it is important to understand variation.

A testing lab wishes to test two brands of paint to see how long each will last before fading. The lab makes 6 gallons of each brand and measures how long it lasts (in months). The data is below.

Brand A	10	60	50	30	40	20
Brand B	35	45	30	35	40	25

Measures of Variation

A testing lab wishes to test two brands of paint to see how long each will last before fading. The lab makes 6 gallons of each brand and measures how long it lasts (in months). The data is below.

Brand A	10	60	50	30	40	20
Brand B	35	45	30	35	40	25

If we compute the mean for each brand, we get 35 months each time (you should verify this).

A testing lab wishes to test two brands of paint to see how long each will last before fading. The lab makes 6 gallons of each brand and measures how long it lasts (in months). The data is below.

Brand A	10	60	50	30	40	20
Brand B	35	45	30	35	40	25

If we compute the mean for each brand, we get 35 months each time (you should verify this).

But Brand B is significantly more consistent than Brand A is (i.e. Brand A has more variation).

Range

The most basic measure of variation of variation is the *range*.

Range

The most basic measure of variation of variation is the *range*.

To find the range of a data set, use the formula $R = max - min$.

Range

The most basic measure of variation of variation is the *range*.

To find the range of a data set, use the formula $R = max - min$.

Graphing calculators will not compute range for us, but they do report the minimum and maximum values.

Range

The number of annual precipitation days for one-half of the 50 largest U.S. cities is listed below. Find the range of the data.

135	128	136	78	116	77	111	79	44	97
116	123	88	102	26	82	156	133	107	35
112	98	45	122	125					

Range

The number of annual precipitation days for one-half of the 50 largest U.S. cities is listed below. Find the range of the data.

135	128	136	78	116	77	111	79	44	97
116	123	88	102	26	82	156	133	107	35
112	98	45	122	125					

By hand: $R = 156 - 26 = 130$

Range

The number of annual precipitation days for one-half of the 50 largest U.S. cities is listed below. Find the range of the data.

135	128	136	78	116	77	111	79	44	97
116	123	88	102	26	82	156	133	107	35
112	98	45	122	125					

By hand: $R = 156 - 26 = 130$

A graphing calculator may make it easier for us to find the minimum and maximum values here.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

We call this difference a deviation.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

We call this difference a deviation.

Note: If we add all the deviations of each data value in a data set, we will always get 0.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

We call this difference a deviation.

Note: If we add all the deviations of each data value in a data set, we will always get 0.

To eliminate this issue, we square the deviations.

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

We call this difference a deviation.

Note: If we add all the deviations of each data value in a data set, we will always get 0.

To eliminate this issue, we square the deviations.

We then find the mean of the squares of the deviations (i.e. add them up and divide by the number of data values).

Variance and Standard Deviation

Before talking about variance and standard deviation, it's important to really understand variation.

Data variation is based on the distance each value is from the mean.

We call this difference a deviation.

Note: If we add all the deviations of each data value in a data set, we will always get 0.

To eliminate this issue, we square the deviations.

We then find the mean of the squares of the deviations (i.e. add them up and divide by the number of data values).

This average is called the variance. Since variance is given in square units, we take its square root and call this the standard deviation.

Variance and Standard Deviation

Below are the formulas for the variance and standard deviation for a *population*. Computing these values by hand is rather time-consuming if we don't use a graphing calculator to compute them for us.

Variance and Standard Deviation

Below are the formulas for the variance and standard deviation for a *population*. Computing these values by hand is rather time-consuming if we don't use a graphing calculator to compute them for us.

Population Variance: $\sigma^2 = \dfrac{\sum(X-\mu)^2}{N}$

Population Standard Deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\sum(X-\mu)^2}{N}}$

Variance and Standard Deviation

Below are the formulas for the variance and standard deviation for a *population*. Computing these values by hand is rather time-consuming if we don't use a graphing calculator to compute them for us.

Population Variance: $\sigma^2 = \dfrac{\sum(X - \mu)^2}{N}$

Population Standard Deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\sum(X - \mu)^2}{N}}$

Note: X is an individual value, μ is population mean, N is population size

Variance and Standard Deviation

Find the population variance and standard deviation for Brand A paint from a previous example.

The data values are $10, 60, 50, 30, 40, 20$. We computed a mean of $\mu = 35$.

Variance and Standard Deviation

Find the population variance and standard deviation for Brand A paint from a previous example.

The data values are $10, 60, 50, 30, 40, 20$. We computed a mean of $\mu = 35$.

Using a graphing calculator, we find a population standard deviation $\sigma \approx 17.1$) (rounded). We can find the variance by squaring the standard deviation; $\sigma^2 \approx 291.7$

Variance and Standard Deviation

When computing variance and standard deviation by hand, it is wise to use a table for organization.

Variance and Standard Deviation

When computing variance and standard deviation by hand, it is wise to use a table for organization.

Value X	$X-\mu$	$(X-\mu)^2$
10	-25	625
60	25	625
50	15	225
30	-5	25
40	5	25
20	-15	225
TOTALS	0	1750

Variance and Standard Deviation

When computing variance and standard deviation by hand, it is wise to use a table for organization.

Value X	$X-\mu$	$(X-\mu)^2$
10	-25	625
60	25	625
50	15	225
30	-5	25
40	5	25
20	-15	225
TOTALS	0	1750

So the variance is $\dfrac{1750}{6} \approx 291.7$ and the standard deviation is $\sqrt{\dfrac{1750}{6}} \approx 17.1$.

Variance and Standard Deviation

Find the population variance and standard deviation of the precipitation data below.

135	128	136	78	116	77	111	79	44	97
116	123	88	102	26	82	156	133	107	35
112	98	45	122	125					

Variance and Standard Deviation

Find the population variance and standard deviation of the precipitation data below.

135	128	136	78	116	77	111	79	44	97
116	123	88	102	26	82	156	133	107	35
112	98	45	122	125					

Using a graphing calculator, we find $\sigma \approx 36.1$; squaring, $\sigma \approx 1301.2$

Variance and Standard Deviation

You may expect that the formula for sample variance is $\frac{\sum(X-\overline{X})}{n}$, but this is not the formula that is generally used.

Variance and Standard Deviation

You may expect that the formula for sample variance is $\frac{\sum(X-\overline{X})}{n}$, but this is not the formula that is generally used.

The issue is that, when the population is large and the sample is small, the above formula tends to underestimate the population variance.

Variance and Standard Deviation

You may expect that the formula for sample variance is $\dfrac{\sum(X-\overline{X})}{n}$, but this is not the formula that is generally used.

The issue is that, when the population is large and the sample is small, the above formula tends to underestimate the population variance.

To take care of this and generate an unbiased estimator, we use the formulas below for sample variance and standard deviation.

Variance and Standard Deviation

You may expect that the formula for sample variance is $\frac{\sum(X-\overline{X})}{n}$, but this is not the formula that is generally used.

The issue is that, when the population is large and the sample is small, the above formula tends to underestimate the population variance.

To take care of this and generate an unbiased estimator, we use the formulas below for sample variance and standard deviation.

Sample Variance: $s^2 = \frac{\sum(X-\overline{X})}{n-1}$

Sample Standard Variation: $s = \sqrt{s^2} = \sqrt{\frac{\sum(X-\overline{X})}{n-1}}$

Variance and Standard Deviation

You may expect that the formula for sample variance is $\frac{\sum(X-\overline{X})}{n}$, but this is not the formula that is generally used.

The issue is that, when the population is large and the sample is small, the above formula tends to underestimate the population variance.

To take care of this and generate an unbiased estimator, we use the formulas below for sample variance and standard deviation.

Sample Variance: $s^2 = \frac{\sum(X-\overline{X})}{n-1}$

Sample Standard Variation: $s = \sqrt{s^2} = \sqrt{\frac{\sum(X-\overline{X})}{n-1}}$

Note: Use the sample variance/standard deviation unless you are explicitly told to use the population values.

The data show the number of public laws passed by the U.S. Congress for a sample of recent years. Find the variance and standard deviation for the data.

$283, 394, 383, 580, 498, 460, 377, 482$

Variance and Standard Deviation

The data show the number of public laws passed by the U.S. Congress for a sample of recent years. Find the variance and standard deviation for the data.

$283, 394, 383, 580, 498, 460, 377, 482$

Using a graphing calculator, we find that $s \approx 91.5$; squaring, $s^2 \approx 8373.6$

Variance and Standard Deviation

We can also find the variance and standard deviation for grouped data.

This information is found the exact same way we found the mean for grouped data, if we use a graphing calculator (this is what I'll be demonstrating).

Variance and Standard Deviation

We can also find the variance and standard deviation for grouped data.

This information is found the exact same way we found the mean for grouped data, if we use a graphing calculator (this is what I'll be demonstrating).

If you want/need to do the calculation by hand, the formulas are below.

Variance: $s^2 = \dfrac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}$

Standard Deviation: $s = \sqrt{s^2} = \sqrt{\dfrac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}}$

Variance and Standard Deviation

We can also find the variance and standard deviation for grouped data.

This information is found the exact same way we found the mean for grouped data, if we use a graphing calculator (this is what I'll be demonstrating).

If you want/need to do the calculation by hand, the formulas are below.

Variance: $s^2 = \dfrac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}$

Standard Deviation: $s = \sqrt{s^2} = \sqrt{\dfrac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}}$

The formulas look ugly, but if we use a table to organize things, the computations are not difficult.

Variance and Standard Deviation

The data show the number of murders in 25 selected cities. Find the variance and standard deviation for the data.

Class limits	Frequency
34–96	13
97–159	2
160–222	0
223–285	5
286–348	1
349–411	1
412–474	0
475–537	1
538–600	2

Variance and Standard Deviation

The data show the number of murders in 25 selected cities. Find the variance and standard deviation for the data.

Class limits	Frequency
34–96	13
97–159	2
160–222	0
223–285	5
286–348	1
349–411	1
412–474	0
475–537	1
538–600	2

First, find the midpoints (you should verify): $65, 128, 191, 254, 317, 380, 443, 506, 569$

Now, using a graphing calculator, we find that $s \approx 167.2$. Squaring this result, $s^2 \approx 27941.8$

Variance and Standard Deviation

The data show the number of murders in 25 selected cities. Find the variance and standard deviation for the data.

Class limits	Frequency
34–96	13
97–159	2
160–222	0
223–285	5
286–348	1
349–411	1
412–474	0
475–537	1
538–600	2

First, find the midpoints (you should verify): $65, 128, 191, 254, 317, 380, 443, 506, 569$

Now, using a graphing calculator, we find that $s \approx 167.2$. Squaring this result, $s^2 \approx 27941.8$

On the next slides, I'll demonstrate solving this problem by hand. If you are using

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65		13		
97-159	128		2		
160-222	191		0		
223-285	254		5		
286-348	317		1		
349-411	380		1		
412-474	443		0		
475-537	506		1		
538-600	569		2		
TOTALS					

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13		
97-159	128	16384	2		
160-222	191	36481	0		
223-285	254	64516	5		
286-348	317	100489	1		
349-411	380	144400	1		
412-474	443	196249	0		
475-537	506	256036	1		
538-600	569	323761	2		
TOTALS					

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	
97-159	128	16384	2	256	
160-222	191	36481	0	0	
223-285	254	64516	5	1270	
286-348	317	100489	1	317	
349-411	380	144400	1	380	
412-474	443	196249	0	0	
475-537	506	256036	1	506	
538-600	569	323761	2	1138	
TOTALS					

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	54925
97-159	128	16384	2	256	32768
160-222	191	36481	0	0	0
223-285	254	64516	5	1270	322580
286-348	317	100489	1	317	100489
349-411	380	144400	1	380	144400
412-474	443	196249	0	0	0
475-537	506	256036	1	506	256036
538-600	569	323761	2	1138	647522
TOTALS					

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	54925
97-159	128	16384	2	256	32768
160-222	191	36481	0	0	0
223-285	254	64516	5	1270	322580
286-348	317	100489	1	317	100489
349-411	380	144400	1	380	144400
412-474	443	196249	0	0	0
475-537	506	256036	1	506	256036
538-600	569	323761	2	1138	647522
TOTALS			25		

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	54925
97-159	128	16384	2	256	32768
160-222	191	36481	0	0	0
223-285	254	64516	5	1270	322580
286-348	317	100489	1	317	100489
349-411	380	144400	1	380	144400
412-474	443	196249	0	0	0
475-537	506	256036	1	506	256036
538-600	569	323761	2	1138	647522
TOTALS			25	4712	

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	54925
97-159	128	16384	2	256	32768
160-222	191	36481	0	0	0
223-285	254	64516	5	1270	322580
286-348	317	100489	1	317	100489
349-411	380	144400	1	380	144400
412-474	443	196249	0	0	0
475-537	506	256036	1	506	256036
538-600	569	323761	2	1138	647522
TOTALS			25	4712	1558720

Variance and Standard Deviation

Class	Midpoint	X_m^2	Frequency	$f \cdot X_m$	$f \cdot X_m^2$
34-96	65	4225	13	845	54925
97-159	128	16384	2	256	32768
160-222	191	36481	0	0	0
223-285	254	64516	5	1270	322580
286-348	317	100489	1	317	100489
349-411	380	144400	1	380	144400
412-474	443	196249	0	0	0
475-537	506	256036	1	506	256036
538-600	569	323761	2	1138	647522
TOTALS			25	4712	1558720

Now, we can use the formulas to find the variance and standard deviation.

Variance and Standard Deviation

$$s^2 = \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}$$

Variance and Standard Deviation

$$s^2 = \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)}$$

$$= \frac{25(1558720) - (4712)^2}{25(24)}$$

Variance and Standard Deviation

$$
\begin{aligned}
s^2 &= \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)} \\
&= \frac{25(1558720) - (4712)^2}{25(24)} \\
&= \frac{38968000 - 22202944}{600}
\end{aligned}
$$

Variance and Standard Deviation

$$
\begin{aligned}
s^2 &= \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)} \\
&= \frac{25(1558720) - (4712)^2}{25(24)} \\
&= \frac{38968000 - 22202944}{600} \\
&\approx 27941.8
\end{aligned}
$$

Variance and Standard Deviation

$$
\begin{aligned}
s^2 &= \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)} \\
&= \frac{25(1558720) - (4712)^2}{25(24)} \\
&= \frac{38968000 - 22202944}{600} \\
&\approx 27941.8 \\
s &= \sqrt{s^2}
\end{aligned}
$$

Variance and Standard Deviation

$$
\begin{aligned}
s^2 &= \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)} \\
&= \frac{25(1558720) - (4712)^2}{25(24)} \\
&= \frac{38968000 - 22202944}{600} \\
&\approx 27941.8 \\
s &= \sqrt{s^2} \\
&= \sqrt{27941.76}
\end{aligned}
$$

Variance and Standard Deviation

$$
\begin{aligned}
s^2 &= \frac{n\left(\sum(f \cdot X_m^2)\right) - \left(\sum(f \cdot X_m)\right)^2}{n(n-1)} \\
&= \frac{25(1558720) - (4712)^2}{25(24)} \\
&= \frac{38968000 - 22202944}{600} \\
&\approx 27941.8 \\
s &= \sqrt{s^2} \\
&= \sqrt{27941.76} \\
&\approx 167.2
\end{aligned}
$$

Next Steps

- Complete Assignment 3
- Begin Module 4
  - Read 3-3
  - Watch Video Lesson #7

Thanks for watching!