Sect 12.3 – Measures of Dispersion

Objective #1   Understanding the need for measures of dispersion.

Solve the following:

Ex. 1 The Nailed It on the Head Corporation is testing two machines that recently had been repaired. A sample of six 2-inch nails is taken from each machine. The first machine produces nails with lengths of 2.000 in, 2.032 in, 1.974 in, 2.000 in, 1.993 in, and 2.001 in, while the second machine produces nails with lengths of 2.000 in, 0.006 in, 3.117 in, 3.940 in, 0.937 in, and 2.000 in. Find the mean, median, and mode of the lengths of the nails for each machine.

Solution:

For machine #1, first write in order from smallest to largest:

     1.974 in, 1.993 in, 2.000 in, 2.000 in, 2.001 in, 2.032 in

Mean = (1.974 + 1.993 + 2.000 + 2.000 + 2.001 + 2.032) ÷ 6

     = 12 ÷ 6 = 2.000 in

Median = (2.000 + 2.000) ÷ 2 = 4.000 ÷ 2 = 2.000 in

Mode = 2.000 in
Thus, the Mean 2.000 in, the Median = 2.000 in, and the
Mode = 2.000 in

For machine #2, first write in order from smallest to largest:

     0.006 in, 1.993 in, 2.000 in, 2.000 in, 3.117 in, 2.032 in

Mean = (0.006 + 0.937 + 2.000 + 2.000 + 3.117 + 3.940) ÷ 6

     = 12 ÷ 6 = 2.000 in

Median = (2.000 + 2.000) ÷ 2 = 4.000 ÷ 2 = 2.000 in

Mode = 2.000 in
Thus, the Mean 2.000 in, the Median = 2.000 in, and the
Mode = 2.000 in

Notice that the two machines had the same mean, median, and mode, but clearly there is a big difference between what these two machines were producing. If the tolerance for the nails was ± 0.05 in, all six nails would be acceptable from the first machine, but only two nails would be acceptable from the second machine. We need a number that can represent how spread out the data is for each machine.

In this section, we want to examine measures of dispersion or how far is the data spread out. This is important to determine since when a company is producing items, we want the variation of the item to be as small as possible so that it we fit within the tolerance.

Objective #2: Finding and Understanding the Range.

One crude way of measuring the dispersion of the data is to find the difference between the largest data value and the smallest data value. We call this measurement the range.

Range
Range = largest data value – smallest data value

Solve the following:

Ex. 2 The Nailed It on the Head Corporation is testing two machines that recently had been repaired. A sample of six 2-inch nails is taken from each machine. The first machine produces nails with lengths of 2.000 in, 2.032 in, 1.974 in, 2.000 in, 1.993 in, and 2.001 in, while the second machine produces nails with lengths of 2.000 in, 0.006 in, 3.117 in, 3.940 in, 0.937 in, and 2.000 in. Find the range of the lengths of the nails for each machine.
Solution:
For the first machine,
Range = 2.032 – 1.974 = 0.058 in
For the second machine,
Range = 3.940 – 0.006 = 3.934 in

The problem with the range is that only measures the dispersion of two values. If either of those values is significantly larger or smaller than the rest of the values, the range will give a distorted measure of dispersion. To address that issue, we can use what is called the standard deviation.

Objective #3:   Understanding and Calculating the Standard Deviation.

To get a more accurate measure of dispersion, we can use what is called the **standard deviation**. The standard deviation is more or less the average of the differences between the mean $\bar{x}$ and the data values x. This number gives a much better sense of how the data is dispersed since all of

the data values are used. The only issue that occurs is that some of the differences between the data values and the mean are positive and some are negative. To calculate the spread of the data, we need to use the magnitude of the differences. Thus, after we calculate the differences, we will square the results in order to ensure we are working with quantities greater than or equal to zero. Once we have done that and added the differences together, we will either divide by the number of data values (n) if the data is the whole population or by the quantity of the number of data values minus one (n – 1) if the data represents only a sample of the population. This result is called the **variance** of the data. If we take the square root of the variance, we get the standard deviation.

Standard Deviation for ungrouped data
Let x represent a data value
Let $\overline{x}$ represent the average of the data
Let s represent the standard deviation
Let n represent the number of data values

1)   If the data represents the whole population, then

$$s = \sqrt{\frac{\text{sum of } (x - \overline{x})^2}{n}}$$

2)   If the data represents a sample of a population, then

$$s = \sqrt{\frac{\text{sum of } (x - \overline{x})^2}{n-1}}$$

<u>Steps for calculating the Standard Deviation for ungrouped data</u>:

1)   Calculate the mean of the data $\overline{x}$.
2)   Calculate the differences of the data values and the mean $(x - \overline{x})$.
3)   Square each of the differences calculated in step #2 and add them together.
4)   Divide the sum by n if the data represents the whole population or by (n – 1) if the data represents a sample.
5)   Take the square root of the result from step #4.

<u>Solve the following:</u>
Ex. 3 The Nailed It on the Head Corporation is testing two machines that recently had been repaired. A sample of six 2-inch nails is taken from each machine. The first machine produces nails with lengths of

2.000 in, 2.032 in, 1.974 in, 2.000 in, 1.993 in, and 2.001 in, while the second machine produces nails with lengths of 2.000 in, 0.006 in, 3.117 in, 3.940 in, 0.937 in, and 2.000 in. Find the standard deviation of the lengths of the nails for each machine.

Solution:

First machine:

1)　　From example #1, $\bar{x}$ = 2.
2)　　Calculate $(x - \bar{x})$ for each data value.

Data	$(x - \bar{x})$
2.000	$(2.000 - 2) =$　0.000
2.032	$(2.032 - 2) =$　0.032
1.974	$(1.974 - 2) = -0.026$
2.000	$(2.000 - 2) =$　0.000
1.993	$(1.993 - 2) = -0.007$
2.001	$(2.001 - 2) =$　0.001

3)　　Square each of the $(x - \bar{x})$ and add them together:

Data	$(x - \bar{x})$	$(x - \bar{x})^2$
2.000	$(2.000 - 2) =$　0.000	$(0.000)^2 = 0.000000$
2.032	$(2.032 - 2) =$　0.032	$(0.032)^2 = 0.001024$
1.974	$(1.974 - 2) = -0.026$	$(-0.026)^2 = 0.000676$
2.000	$(2.000 - 2) =$　0.000	$(0.000)^2 = 0.000000$
1.993	$(1.993 - 2) = -0.007$	$(-0.007)^2 = 0.000049$
2.001	$(2.001 - 2) =$　0.001	$(0.001)^2 = 0.000001$

Sum of $(x - \bar{x})^2 = 0.00175$

4) Since this is a sample, divide the sum by (n – 1) = 5:

Variance = $\frac{0.00175}{5}$ = 0.00035

5) Take the square root of the result.

s = $\sqrt{0.00035}$ = 0.01870… ≈ 0.019 in

Second machine:

1) From example #1, $\bar{x}$ = 2.
2) Calculate (x – $\bar{x}$) for each data value.

Data	(x – $\bar{x}$)
2.000	(2.000 – 2) = 0.000
0.006	(0.006 – 2) = – 1.994
3.117	(3.117 – 2) = 1.117
3.940	(3.940 – 2) = 1.940
0.937	(0.937 – 2) = – 1.063
2.000	(2.000 – 2) = 0.000

3) Square each of the (x – $\bar{x}$) and add them together:

Data	(x – $\bar{x}$)	(x – $\bar{x}$)^2
2.000	(2.000 – 2) = 0.000	$(0.000)^2$ = 0.000000
0.006	(0.006 – 2) = – 1.994	$(-1.994)^2$ = 3.976036
3.117	(3.117 – 2) = 1.117	$(1.117)^2$ = 1.247689
3.940	(3.940 – 2) = 1.940	$(1.940)^2$ = 3.763600
0.937	(0.937 – 2) = – 1.063	$(-1.063)^2$ = 1.129969
2.000	(2.000 – 2) = 0.000	$(0.000)^2$ = 0.000000

Sum of (x – $\bar{x}$)^2 = 10.117294

4)   Since this is a sample, divide the sum by (n – 1) = 5:
      Variance = $\dfrac{10.117294}{5}$ = 2.0234588

5)   Take the square root of the result.
      s = $\sqrt{2.0234588}$ = 1.4224… ≈ 1.422 in

Ex. 4 During an outbreak of strep throat, a local elementary school
       recorded the number of students out sick with strep throat on a
       particular day for the first two full weeks of December. The data was
       as follows:

| 25 | 34 | 53 | 61 | 77 |
| 52 | 53 | 45 | 24 | 13 |

Find the standard deviation.

1) $\bar{x}$ = (25 + 34 + 53 + 61 + 77 + 52 + 53 + 45 + 24 + 13) ÷ 10

      = 437 ÷ 10 = 43.7 students per day.

2)   Calculate (x – $\bar{x}$) for each data value.

Data	(x – $\bar{x}$)
25	(25 – 43.7) = – 18.7
34	(34 – 43.7) = – 9.7
53	(53 – 43.7) = 9.3
61	(61 – 43.7) = 17.3
77	(77 – 43.7) = 33.3
52	(52 – 43.7) = 8.3
53	(53 – 43.7) = 9.3
45	(45 – 43.7) = 1.3
24	(24 – 43.7) = – 19.7
13	(13 – 43.7) = – 30.7

3)   Square each of the $(x - \bar{x})$ and add them together:

Data	$(x - \bar{x})$	$(x - \bar{x})^2$
25	$(25 - 43.7) = -18.7$	$(-18.7)^2 = 349.69$
34	$(34 - 43.7) = -9.7$	$(-9.7)^2 = 94.09$
53	$(53 - 43.7) = 9.3$	$(9.3)^2 = 86.49$
61	$(61 - 43.7) = 17.3$	$(17.3)^2 = 299.29$
77	$(77 - 43.7) = 33.3$	$(33.3)^2 = 1108.89$
52	$(52 - 43.7) = 8.3$	$(8.3)^2 = 68.89$
53	$(53 - 43.7) = 9.3$	$(9.3)^2 = 86.49$
45	$(45 - 43.7) = 1.3$	$(1.3)^2 = 1.69$
24	$(24 - 43.7) = -19.7$	$(-19.7)^2 = 388.09$
13	$(13 - 43.7) = -30.7$	$(-30.7)^2 = 942.49$

Sum of $(x - \bar{x})^2 = 3426.1$

4)   Since this is not a sample, divide the sum by n = 10:
Variance = $\frac{3426.1}{10}$ = 342.61

5)   Take the square root of the result.
s = $\sqrt{342.61}$ = 18.509… ≈ 18.5 students

To calculate the mean and standard deviation on a scientific calculator, we need to first enter the data into a calculator. To do this on a TI 30-Xa, enter the first data value and Σ+ key (directly above the STO key). The display should read n = 1. You should also see STAT at the top of the screen. Repeat this procedure for all of the other data values. Also, you enter more data values, the value of n goes up. Once you are finished, the value of n should be equal to the number of values in your data. If not, then you made a mistake and need to start over. To do this, hit CSR key (2nd and then 7

on a TI-30Xa). To find the mean, hit the $\bar{x}$ key (2nd and then x^2). To find the standard deviation for a sample, hit the σxn-1 key (2nd and then $\sqrt{x}$). To find the standard deviation for a sample, hit the σxn key (2nd and then ÷). You can even find out what the sum of $(x - \bar{x})^2$ by hitting the $\sum x^2$ key (2nd and then)). To get your calculator out of STAT mode, hit the CSR key (2nd and then 7 or turn it off and then back on). As an exercise, check the mean and standard deviation for the last example on your calculator. Note, the scientific calculatro does not allow for the calculation of the grouped mean or the grouped standard deviation.

Objective #4   Understanding and Calculating the Standard Deviation for grouped data.

To calculate the standard deviation for grouped data, instead of calculate the difference between the data value and the mean and squaring the result, we will need to calculate the difference between the midpoint and the mean for each class and squaring the result and multiply it by the appropriate frequency. We then proceed as before.

Standard Deviation for grouped data
Let M represent a midpoint of a class interval
Let F represent the frequency
Let $\bar{x}$ represent the average of the data
Let s represent the standard deviation
Let n represent the number of data values

1)   If the data represents the whole population, then

$$s = \sqrt{\frac{\text{sum of } F(M-\bar{x})^2}{n}}$$

2)   If the data represents a sample of a population, then

$$s = \sqrt{\frac{\text{sum of } F(M-\bar{x})^2}{n-1}}$$

Steps for calculating the Standard Deviation for grouped data:

1) Calculate the grouped mean of the data $\bar{x}$.
2) For each class interval, calculate the differences of the mipoint and the mean $(M - \bar{x})$.
3) For each class interval, square the differences calculated in step #2 and multiply by the frequency $F(M - \bar{x})^2$. Add the results together.
4) Divide the sum by n if the data represents the whole population or by $(n - 1)$ if the data represents a sample.
5) Take the square root of the result from step #4

Solve the following:

Ex. 5 A large company gathered data on the length of time each of its employees has been employed by the company. The results are summarized below:

Years of Service	Frequency F	Years of Service	Frequency F
0 – 6	145	28 – 34	33
7 – 13	409	35 – 41	16
14 – 20	231	42 – 48	5
21 – 27	160	49 – 55	1

Calculate the grouped standard deviation.
Solution:
1)   a)   We first need to calculate the midpoint of each class:

(6 + 0)/2 = 3   (7 + 13)/2 = 10   (14 + 20)/2 = 17
(21 + 27)/2 = 24   (28 + 34)/2 = 31   (35 + 41)/2 = 38
(42 + 48)/2 = 45   (49 + 55)/2 = 52

Years of Service	Frequency F	Midpoint M
0 – 6	145	3
7 – 13	409	10
14 – 20	231	17
21 – 27	160	24
28 – 34	33	31
35 – 41	16	38
42 – 48	5	45
49 – 55	1	52

b)   Next, we need to calculate the FM products:

145•3 = 435   357•10 = 3570   231•17 = 3927
162•24 = 3888   33•31 = 1023   16•38 = 608
5•45 = 225   1•52 = 52

Years of Service	Frequency F	Midpoint M	FM Products
0 – 6	145	3	435
7 – 13	409	10	4090
14 – 20	231	17	3927
21 – 27	160	24	3840
28 – 34	33	31	1023
35 – 41	16	38	608
42 – 48	5	45	225
49 – 55	1	52	52

c)   Now, we add the frequencies to find n and add the FM products:

Years of Service	Frequency F	Midpoint M	FM Products
0 – 6	145	3	435
7 – 13	409	10	4090
14 – 20	231	17	3927
21 – 27	160	24	3840
28 – 34	33	31	1023
35 – 41	16	38	608
42 – 48	5	45	225
49 – 55	1	52	52

n = 1000   Sum FM = 14200

d)   Lastly, divide Sum FM by n to find the grouped mean
$\overline{x}$ = 14200/1000 = 14.2

Thus, the grouped mean is 14.2 years.

2)   For each class interval, we will calculate $(M - \overline{x})$.
(3 – 14.2) = – 11.2   (10 – 14.2) = – 4.2   (17 – 14.2) = 2.8
(24 – 14.2) = 9.8   (31 – 14.2) = 16.8   (38 – 14.2) = 23.8
(45 – 14.2) = 30.8   (52 – 14.2) = 37.8

Years of Service	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$
0 – 6	145	3	435	– 11.2
7 – 13	409	10	4090	– 4.2
14 – 20	231	17	3927	2.8
21 – 27	160	24	3840	9.8
28 – 34	33	31	1023	16.8
35 – 41	16	38	608	23.8
42 – 48	5	45	225	30.8
49 – 55	1	52	52	37.8

3) Next, for each class interval, we will square $(M - \bar{x})$:

$(-11.2)^2 = 125.44$ $\quad$ $(-4.2)^2 = 17.64$ $\quad$ $(2.8)^2 = 7.84$

$(9.8)^2 = 96.04$ $\quad$ $(16.8)^2 = 282.24$ $\quad$ $(23.8)^2 = 566.44$

$(30.8)^2 = 948.64$ $\quad$ $(37.8)^2 = 1428.84$

Years of Service	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$	$(M - \bar{x})^2$
0 – 6	145	3	435	– 11.2	125.44
7 – 13	409	10	4090	– 4.2	17.64
14 – 20	231	17	3927	2.8	7.84
21 – 27	160	24	3840	9.8	96.04
28 – 34	33	31	1023	16.8	282.24
35 – 41	16	38	608	23.8	566.44
42 – 48	5	45	225	30.8	948.64
49 – 55	1	52	52	37.8	1,428.84

Now, for each class interval, multiply $(M - \bar{x})^2$ by F:

145(125.44) = 18,188.8 $\qquad$ 409(17.64) = 7,214.76

231(7.84) = 1,811.04 $\qquad$ 160(96.04) = 15,366.4

33(282.24) = 9,313.92 $\qquad$ 16(566.44) = 9,063.04

5(948.64) = 4,743.2 $\qquad$ 1(1,428.84) = 1,428.84

Then, add up the results:

Years of Service	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$	$(M - \bar{x})^2$	$F(M - \bar{x})^2$
0 – 6	145	3	435	– 11.2	125.44	18,188.8
7 – 13	409	10	4090	– 4.2	17.64	7,214.76
14 – 20	231	17	3927	2.8	7.84	1,811.04
21 – 27	160	24	3840	9.8	96.04	15,366.4
28 – 34	33	31	1023	16.8	282.24	9,313.92
35 – 41	16	38	608	23.8	566.44	9,063.04
42 – 48	5	45	225	30.8	948.64	4,743.2
49 – 55	1	52	52	37.8	1,428.84	1,428.84

n = 1000   Sum of $F(M - \bar{x})^2$ = 67,130

4) Divide the sum by 1000 since the data represents the whole population
Variance = 67130/1000 = 67.13

5) Take the square root of the result to get the standard deviation:
s = $\sqrt{67.13}$ = 8.19… ≈ 8.2 years.

Ex. 6 In a random sample of their customers, PetBrain asked their customers how many different months out of the past year did they make a purchase from PetBrain. The results are given below:

Number of Months	Frequency F
1 – 3	21
4 – 6	86
7 – 9	98
10 – 12	45

Calculate the grouped standard deviation.
Solution:
1) a) We first need to calculate the midpoint of each class:
(1 + 3)/2 = 2   (4 + 6)/2 = 5   (7 + 9)/2 = 8
(10 + 12)/2 = 11

Number of Months	Frequency F	Midpoint M
1 – 3	21	2
4 – 6	86	5
7 – 9	98	8
10 – 12	45	11

b)   Next, we need to calculate the FM products:

21•2 = 42   86•5 = 430   98•8 = 784

45•11 = 495

Number of Months	Frequency F	Midpoint M	FM Products
1 – 3	21	2	42
4 – 6	86	5	430
7 – 9	98	8	784
10 – 12	45	11	495

c)   Now, we add the frequencies to find n and add the FM products:

Number of Months	Frequency F	Midpoint M	FM Products
1 – 3	21	2	42
4 – 6	86	5	430
7 – 9	98	8	784
10 – 12	45	11	495

n = 250   Sum FM = 1751

d)   Lastly, divide Sum FM by n to find the grouped mean
$\overline{x}$ = 1751/250 = 7.004

Thus, the grouped mean is 7.004 different months.

2)   For each class interval, we will calculate (M – $\overline{x}$).

(2 – 7.004) = – 5.004   (5 – 7.004) = – 2.004
(8 – 7.004) = 0.996   (11 – 7.004) = 3.996

Number of Months	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$
1 – 3	21	2	42	– 5.004
4 – 6	86	5	430	– 2.004
7 – 9	98	8	784	0.996
10 – 12	45	11	495	3.996

3) Next, for each class interval, we will square $(M - \bar{x})$:

$(-5.004)^2 = 25.040016$ $(-2.004)^2 = 4.016016$

$(0.996)^2 = 0.992016$ $(3.996)^2 = 15.968016$

# of Months	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$	$(M - \bar{x})^2$
1 – 3	21	2	42	– 5.004	25.040016
4 – 6	86	5	430	– 2.004	4.016016
7 – 9	98	8	784	0.996	0.992016
10 – 12	45	11	495	3.996	15.968016

Now, for each class interval, multiply $(M - \bar{x})^2$ by F:

21(25.040016) = 525.840336   86(4.016016) = 345.377376

981(0.992016) = 97.217568   45(15.968016) = 718.560720

Then, add up the results:

# of Months	Frequency F	Midpoint M	FM Products	$(M - \bar{x})$	$(M - \bar{x})^2$	$F(M - \bar{x})^2$
1 – 3	21	2	42	– 5.004	25.040016	525.840336
4 – 6	86	5	430	– 2.004	4.016016	345.377376
7 – 9	98	8	784	0.996	0.992016	97.217568
10 – 12	45	11	495	3.996	15.968016	718.560720

n = 250   Sum of $F(M - \bar{x})^2$ = 1,686.996

4) Divide the sum by 250 – 1 = 249 since the data represents the sample population:
Variance = 1686.996/249 = 6.775…

5) Take the square root of the result to get the standard deviation:
s = $\sqrt{6.775...}$ = 2.60… ≈ 2.6 months.