2.4 Measures of Variation

Range

A range of a data set is the difference between the maximum and minimum data entries in the set. To find the range, the data must be quantitative.

Range = largest - smallest

Example 1

Two corporations each hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries for Corporation A, then find the range of the starting salaries for Corporations B. Compare the two corporations.

Starting Salaries for Corporation A (in thousands of dollars)

Salary	41	38	39	45	47	41	44	41	37	42

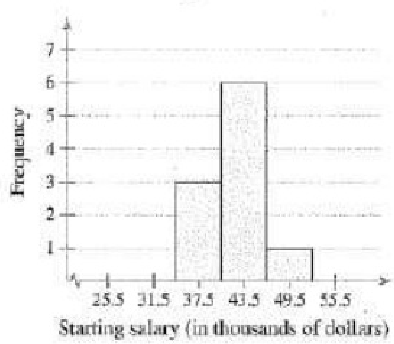
47-37 = 10

Starting Salaries for Corporation B (in thousands of dollars)

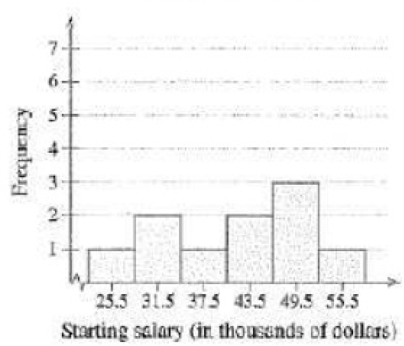
Salary	40	23	41	50	49	32	41	29	52	58

58 - 23 = 35

Corporation A



Corporation B



Variance and Standard Deviation

The deviation of an entry x in a population data set is the difference between the entry and the mean μ of the data set.

deviation of $x = x - \mu$

The population variance of a population data set of N entries is

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$$

SS_x Sum of the Squares

"sigma"

The symbol σ is the lowercase Greek letter sigma.

The population standard deviation of a population data set of N entries is the square root of the population variance. $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

- The standard deviation measures the variation of the data set about the mean and has the same units of measure as the data set.
- The standard deviation is always greater than or equal to 0. When $\sigma = 0$, the data set has no variation and all entries have the same value.
- As the entries get farther from the mean (that is, more spread out), the value of σ increases.

Finding the Population Variance and Standard Deviation

1. Find the mean of the population data set. $\mu = \frac{\Sigma x}{N}$
2. Find the deviation of each entry. $x - \mu$
3. Square each deviation. $(x-\mu)^2$
4. Add to get the sum of squares. $SS_x = \Sigma(x-\mu)^2$
5. Divide by N to get the population variance. $\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$
6. Find the square root of the variance to get the population standard deviation.

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

Example 2

Find the population variance and standard deviation of the starting salaries for Corporation B in example 1.

$\mu = 41.5$

Salary	40	23	41	50	49	32	41	29	52	58
$X-\mu$	$40-41.5 = -1.5$	-18.5	$-.5$	8.5	7.5	-9.5	$-.5$	-12.5	10.5	16.5
$(X-\mu)^2$	2.25	342.25	.25	72.25	56.25	90.25	.25	156.25	110.25	272.25

$SS_x = 1102.5$ $\quad \sigma^2 = \frac{1102.5}{10} = 110.25$

$\sigma = \sqrt{110.25} = 10.5$

The sample variance and sample standard deviation of a sample data set of n entries are:

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

The process is the same, but you use $\bar{X}$ in place of μ and you use $n-1$ in place of N.

$s \to \sigma$

Example 3

In a study of high school football players that suffered concussions, researchers placed the players in two groups. Players that recovered from their concussions in 14 days or less were placed in Group 1. Those that took more than 14 days were place in Group 2. The recovery times (in days) for Group 2 are listed below. Find the sample variance and the standard deviation of the recovery times.

43, 57, 18, 45, 47, 33, 49, 24

Calc output

$\boxed{\bar{x} = 39.5}$

$\sum x = 319$

$\sum x^2 = 13722$

$\boxed{S_x = 13.309}$ Sample st. dev.

$\sigma_x = 12.45$

$n = 8$

variance $S^2 \approx 177.1$

x	$x-\bar{x}$	$(x-\bar{x})^2$
43	$43-39.5$	
57	$57-39.5$	
18	⋮	
45		
47		
33		
49		
24		

$$\sqrt{\frac{\sum (x-\bar{x})^2}{7}}$$

To put in a list: $\boxed{\text{STAT}}$ $\boxed{1}$ EDIT

to clear: ^ List1 $\boxed{\text{Clear}}$ $\boxed{\text{Enter}}$

To find stats: $\boxed{\text{STAT}}$ $\boxed{\rightarrow}$ CALC $\boxed{1}$ 1-Var stats or $\boxed{\text{ENTER}}$

newer O.S. List: $\boxed{L_1}$ Freq List: (leave blank) Calculate $\boxed{\text{ENTER}}$

Using Technology

Example 4

Sample office rental rates (in dollars per square foot per year) for Los Angeles are shown in the table. Use technology to find the mean rental rate and the sample standard deviation.

Office rental rates		
69	29	46
24	18	43
20	25	19
24	22	35
24	28	32
30	29	20
25	38	27
60	25	31

MINITAB

Descriptive Statistics: Rental Rates

Variable	N	Mean	SE Mean	StDev	Minimum
Rental Rates	24	30.96	2.57	12.59	18.00

Variable	Q1	Median	Q3	Maximum
Rental Rates	24.00	27.50	34.25	69.00

EXCEL

	A	B
1	Mean	30.95833
2	Standard Error	2.569666
3	Median	27.5
4	Mode	24
5	Standard Deviation	12.58874
6	Sample Variance	158.4764
7	Kurtosis	3.255136
8	Skewness	1.809862
9	Range	51
10	Minimum	18
11	Maximum	69
12	Sum	743
13	Count	24

TI-84 PLUS

1-Var Stats
$\bar{x}$=30.95833333
Σx=743
Σx^2=26647
Sx=12.58874296
σx=12.32368711
↓n=24

Sample Mean
Sample Standard Deviation

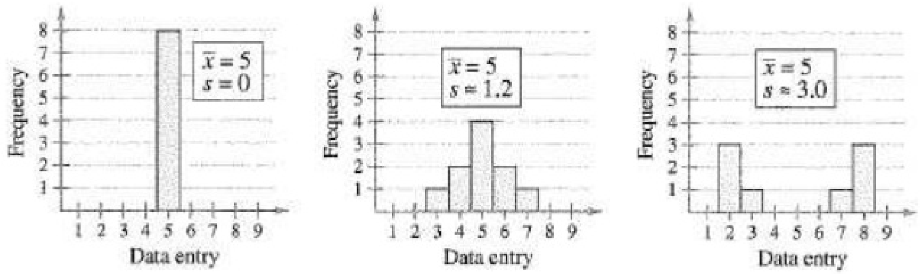
Sample office rental rates (in dollars per square foot per year) for the Dallas Fort Worth area are shown in the table. Use technology to find the mean rental rate and the sample standard deviation.

Office rental rates		
22	35	18
21	27	16
18	22	16
24	20	17
15	31	24
25	24	23

1-var stats L_2 $\bar{x} = 22.1$

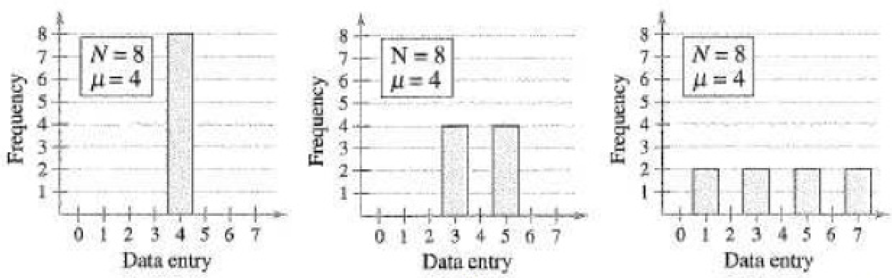
$S_x = 5.3$

Interpreting Standard Deviation



Example 5

Without calculating, estimate the population standard deviation of each data set.



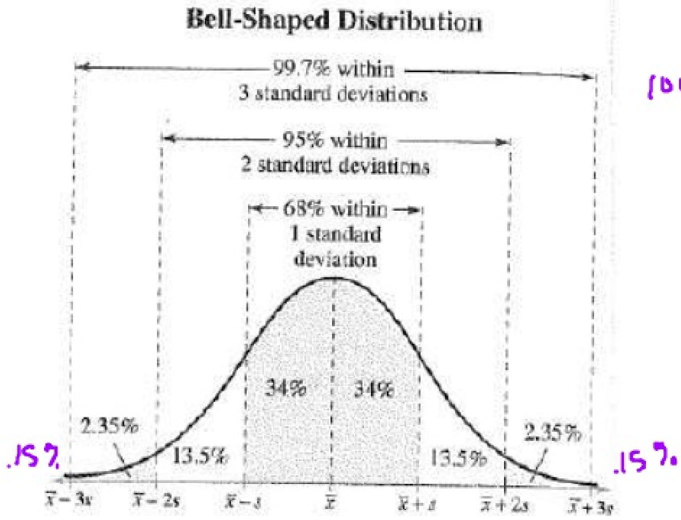
$\sigma = 0$ $\sigma = 1$ $\sigma \approx 2.2$

Data entries that lie more than two standard deviations from the mean are considered unusual, while those that lie more than three standard deviations from the mean are very unusual.

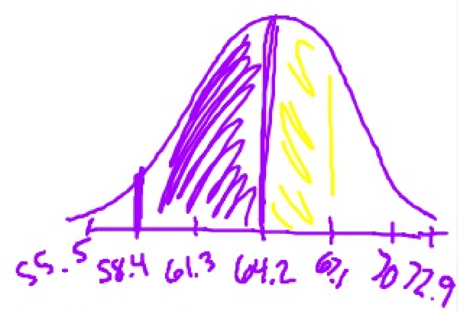
Empirical Rule or 68 – 95 – 99.7 Rule

For data sets with distributions that are approximately symmetric and bell-shaped, the standard deviation has these characteristics.

1. About 68 of the data lie within 1 standard deviations of the mean.
2. About 95 of the data lie within 2 standard deviations of the mean.
3. About 99.7 of the data lie within 3 standard deviations of the mean.



100 − 99.7 = .3



Example 6

In a survey conducted by the National center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64.2 inches, with a sample standard deviation of 2.9 inches. Estimate the percent of women whose heights are between 58.4 and 64.2 inches.

$\frac{95\%}{2}$ or 34% + 13.5%

$\boxed{47.5\%}$

Estimate the percent of women ages 20-29 whose heights are between 64.2 inches and 67.1 inches. 34%

Chebychev's Theorem

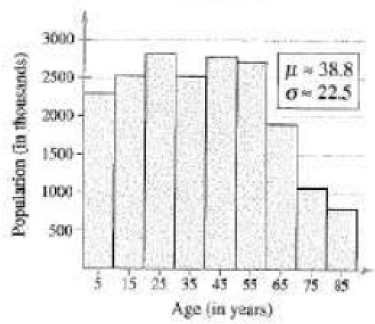
The portion of any data set lying within k standard deviations ($k > 1$) of the mean is at least $1 - \frac{1}{K^2}$

- $k = 2$: In any data set, at least $1-\frac{1}{2^2} = 1-\frac{1}{4}$, or $3/4$ (75%) of the data lie within 2 standard deviations of the mean.
- $k = 3$: In any data set, at least $1-\frac{1}{3^2} = 1-\frac{1}{9}$, or $8/9$ (88.9%) of the data lie within 3 standard deviations of the mean.

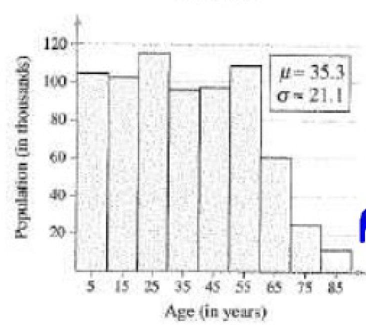
Example 7

The age distributions for New York and Alaska are shown in the histograms. Apply Chebychev's Theorem to the data for New Your using $k = 2$.

New York



Alaska



$\mu - 2\sigma$

$38.8 - 2(22.5) = -6.2$

$38.8 + 2(22.5) = 83.8$

At least 75% of the population in NY is between 0 and 83.8

Apply Chebychev's Theorem to the data for Alaska using $k = 2$.

$35.3 - 2(21.1) = -6.9$

$35.3 + 2(21.1) = 77.5$

At least 75% of the pop. in Alaska is between 0 and 77.5

Standard Deviation for Grouped Data

Sample standard deviation = $s = \sqrt{\frac{\Sigma(x-\bar{x})^2 \cdot f}{n-1}}$

where $n = \Sigma f$ is the number of entries in the data set.

Example 8

You collect a random sample of the number of children per household in a region. The results are shown in the table. Find the sample mean and the sample standard deviation of the data set.

Number of children in 50 households				
1	3	1	1	1
1	2	2	1	0
1	1	0	0	0
1	5	0	3	6
3	0	3	1	1
1	1	6	0	1
3	6	6	1	2
2	3	0	1	1
4	1	1	2	2
0	3	0	2	4

x	f	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
0	10	-1.82	3.3124	33.1
1	19	-.82	0.6724	12.8
2	7	.18	0.0324	0.2
3	7	1.18	1.3924	9.7
4	2 5	2.18	4.7524	9.5
5	1	3.18	10.1124	10.1
6	4 1	4.18	17.4724	69.9
	$\Sigma f = 50$			

$$\frac{\Sigma (x-\bar{x})^2 f}{n-1} = \frac{145.3}{49}$$

$$\bar{x} = \frac{0+19+14+21+8+5+24}{50} = 1.82$$

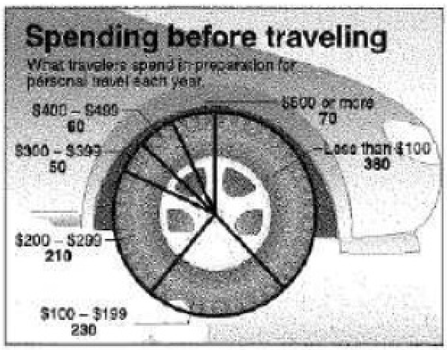
$$s = \sqrt{\frac{145.3}{49}} \approx 1.7$$

Change three of the 6's in the data set to 4's. How does this change affect the sample mean and sample standard deviation?

$\bar{x} = 1.7$ $\quad$ $s = 1.5$

Example 9

The figure at the right shows the results of a survey in which 1000 adults were asked how much they spend in preparation for personal travel each year. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set.



Class	x	f	xf	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
0–99	49.5	380	18,810	−142.5	20,306.25	7,716,375.0
100–199	149.5	230	34,385	−42.5	1,806.25	415,437.5
200–299	249.5	210	52,395	57.5	3,306.25	694,312.5
300–399	349.5	50	17,475	157.5	24,806.25	1,240,312.5
400–499	449.5	60	26,970	257.5	66,306.25	3,978,375.0
500+	599.5	70	41,965	407.5	166,056.25	11,623,937.5
		$\Sigma = 1000$	$\Sigma = 192{,}000$			$\Sigma = 25{,}668{,}750.0$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{192{,}000}{1000} = 192$$ Sample mean

Use the sum of squares to find the sample standard deviation.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{25{,}668{,}750}{999}} \approx 160.3$$ Sample standard deviation

So, the sample mean is $192 per year, and the sample standard deviation is about $160.30 per year.

Coefficient of Variation

The coefficient of variation (CV) of a data set describes the standard deviation as a percent of the mean.

Population: $CV = \frac{\sigma}{\mu} \cdot 100\%$ Sample: $CV = \frac{s}{\bar{x}} \cdot 100\%$

Example 10

The table shows the population heights (in inches) and weights (in pounds) of the members of a basketball team. Find the coefficient of variation for the heights and the weights. Then compare the results.

Heights and Weights of a Basketball Team

Heights	Weights
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

Heights

$\mu = \bar{x} = 72.75$

$\sigma = 3.29$

$\frac{3.29}{72.75} \times 100$

$\boxed{4.5\%}$

Weights

$\mu = \bar{x} = 187.83$

$\sigma = 17.69$

$\boxed{9.4\%}$

Find the coefficient of variation for the office rental rates in Los Angeles (example 4) and for those in the Dallas/Fort Worth area. Then compare the results.