NUMERICAL MEASURES

Prepared by: Jasmine A. Tulin

MEASURES OF LOCATION

Prepared by: Jasmine A. Tulin

ARITHMETIC MEAN

Mean/Arithmetic mean

Average value for a variable.

The mean is denoted by $\bar{x}$.

n = sample size

x_1 = value of variable x for the first observation

x_2 = value of variable x for the second observation

x_n = value of variable x for the nth observation

Sample mean, $\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdot \cdot \cdot + x_n}{n}$

Table 2.9 - Data on Home Sales in Cincinnati, Ohio, Suburb

- Illustration: Computation of the mean home selling price for the sample of 12 home sales:

Home Sale	Selling Price ($)
1	138,000
2	254,000
3	186,000
4	257,500
5	108,000
6	254,000
7	138,000
8	298,000
9	199,500
10	208,000
11	142,000
12	456,250

MEDIAN

- **Median**: Value in the middle when the data are arranged in ascending order.
  - Middle value, for an odd number of observations
  - Average of two middle values, for an even number of observations

MODE

Mode: Value that occurs most frequently in a data set.

Consider the class size data:

32 42 46 46 54

Observe - 46 is the only value that occurs more than once.

Mode is 46.

Multimodal data - Data contain at least two modes.

Bimodal data - Data contain exactly two modes.

Calculating the Mean, Median, and Modes for the Home Sales Data using Excel

	A	B	C	D	E
1	Home Sale	Selling Price ($)			
2	1	138,000		Mean:	=AVERAGE(B2:B13)
3	2	254,000		Median:	=MEDIAN(B2:B13)
4	3	186,000		Mode 1:	=MODE.MULT(B2:B13)
5	4	257,500		Mode 2:	=MODE.MULT(B2:B13)
6	5	108,000			
7	6	254,000			
8	7	138,000			
9	8	298,000			
10	9	199,500			
11	10	208,000			
12	11	142,000			
13	12	456,250			

	A	B	C	D	E
1	Home Sale	Selling Price ($)			
2	1	138,000		Mean:	$ 219,937.50
3	2	254,000		Median:	$ 203,750.00
4	3	186,000		Mode 1:	$ 138,000.00
5	4	257,500		Mode 2:	$ 254,000.00
6	5	108,000			
7	6	254,000			
8	7	138,000			
9	8	298,000			
10	9	199,500			
11	10	208,000			
12	11	142,000			
13	12	456,250			

Measures of VARIABILITY

Prepared by: Jasmine A. Tulin

RANGE

Range: Found by subtracting the smallest value from the largest value in a data set.

Illustration: Consider the data on home sales in Cincinnati, Ohio, Suburb:

Home Sale	Selling Price ($)
1	138,000
2	254,000
3	186,000
4	257,500
5	108,000
6	254,000
7	138,000
8	298,000
9	199,500
10	208,000
11	142,000
12	456,250

VARIANCE

Variance: Measure of variability that utilizes all the data.

It is based on the deviation about the mean, which is the difference between the value of each observation (x_i) and the mean.

The deviations about the mean are squared while computing the variance.

Sample variance, $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

Population variance , $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$

Computation of Deviations and Squared Deviations about the Mean for the Class Size Data

Number of Students in Class (x_i)	Mean Class Size ($\bar{x}$)	Deviation About the Mean ($x_i - \bar{x}$)	Squared Deviation About the Mean $(x_i - \bar{x})^2$
46	44	2	4
54	44	10	100
42	44	−2	4
46	44	2	4
32	44	−12	144
		0	256
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

Computation of Sample Variance:

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{256}{4} = 64$$

STANDARD DEVIATION; COEFFICIENT OF VARIATION

Standard deviation: Positive square root of the variance

Measured in the same units as the original data.

For sample , $s = \sqrt{s^2}$

For population, $\sigma = \sqrt{\sigma^2}$

Coefficient of variation:

$$\left(\frac{\text{Standard deviation}}{\text{Mean}} \text{ x } 100\right)\%$$

Measures the standard deviation relative to the mean.

Expressed as a percentage.

Computation of Coefficient of Variation

Illustration:

Consider the class size data:

46 54 42 46 32

Mean, $\bar{x} = 44$

Standard deviation, $s = 8$

Coefficient of variation $= \left(\frac{8}{44} \text{ x } 100\right)\% = 18.2\%$

Calculating Variability Measures for the Home Sales Data in Excel

	A	B	C	D	E
1	Home Sale	Selling Price ($)			
2	1	138000		Mean:	=AVERAGE(B2:B13)
3	2	254000		Median:	=MEDIAN(B2:B13)
4	3	186000		Mode 1:	=MODE.MULT(B2:B13)
5	4	257500		Mode 2:	=MODE.MULT(B2:B13)
6	5	108000			
7	6	254000		Range:	=MAX(B2:B13)-MIN(B2:B13)
8	7	138000		Variance:	=VAR.S(B2:B13)
9	8	298000		Standard Deviation:	=STDEV.S(B2:B13)
10	9	199500			
11	10	208000		Coefficient of Variation:	=E9/E2
12	11	142000			
13	12	456250		85th Percentile:	=PERCENTILE.EXC(B2:B13,0.85)

	A	B	C	D	E
1	Home Sale	Selling Price ($)			
2	1	138,000		Mean:	$ 219,937.50
3	2	254,000		Median:	$ 203,750.00
4	3	186,000		Mode 1:	$ 138,000.00
5	4	257,500		Mode 2:	$ 254,000.00
6	5	108,000			
7	6	254,000		Range:	$ 348,250.00
8	7	138,000		Variance:	9037501420
9	8	298,000		Standard Deviation:	$ 95,065.77
10	9	199,500			
11	10	208,000		Coefficient of Variation:	43.22%
12	11	142,000			
13	12	456,250		85th Percentile:	$ 305,912.50

ANALYZING DISTRIBUTIONS

Prepared by: Jasmine A. Tulin

PERCENTILE

- **Percentile**: Value of a variable at which a specified (approximate) percentage of observations are below that value.
  - The *p*th percentile tells us the point in the data where:
    - Approximately *p* percent of the observations have values less than the *p*th percentile;
    - Approximately (100 – *p*) percent of the observations have values greater than the *p*th percentile.

PERCENTILE

- Steps to calculate the pth percentile:
  - Arrange the data in ascending order (smallest to largest value).
  - Compute $k = (n + 1) \times p$.
  - Divide k into its integer component, i, and its decimal component, d.
    a. If $d = 0$, find the kth largest value in the data set. This is the pth percentile.

(*contd.*)

PERCENTILE

b. If $d > 0$, the percentile is between the values in positions i and $i + 1$ in the sorted data. To find this percentile, we must interpolate between these two values.

   i. Calculate the difference between the values in positions i and $i + 1$ in the sorted data set. We define this difference between the two values as m.

   ii. Multiply this difference by d: $t = m \times d$.

   iii. To find the pth percentile, add t to the value in position i of the sorted data.

PERCENTILE

- Illustration: To determine the 85th percentile for the home sales data in Table 2.9.
- 1. Arrange the data in ascending order.
- 108,000 138,000 138,000 142,000 186,000 199,500 208,000 254,000 254,000 257,500 298,000 456,250
- 2. Compute $k = (n + 1) \times p = (12 + 1) \times 0.85 = 11.05$.
- 3. Dividing 11.05 into the integer and decimal components gives us $i = 11$ and $d = 0.05$.
  - $d > 0$, interpolate between the values in the 11th and 12th positions in the sorted data.

PERCENTILE

- Illustration (contd.): To determine the 85th percentile for the home sales data in Table 2.9.
  - The value in the 11th position is 298,000, and
  - The value in the 12th position is 456,250.
- i. $m = 456{,}250 - 298{,}000 = 158{,}250$
- ii. $t = m \times d = 158{,}250 \times 0.05 = 7912.5$
- iii. pth percentile = $298{,}000 + 7912.5 = 305{,}912.5$
- $305,912.50 represents the 85th percentile of the home sales data.

QUARTILES

- **Quartiles**:
  - When the data is divided into four equal parts:
    - Each part contains approximately 25% of the observations.
    - Division points are referred to as quartiles.

Q_1 = first quartile, or 25th percentile

Q_2 = second quartile, or 50th percentile (also the median)

Q_3 = third quartile, or 75th percentile

Z-SCORE

- ***z*-score**:
  - Measures the relative location of a value in the data set.
  - Helps to determine how far a particular value is from the mean relative to the data set's standard deviation.
  - Standardized value
- If $x_1, x_2, . . . , x_n$ is a sample of *n* observations
  - $z_i = \frac{x_i - \bar{x}}{s}$
    - z_i = *z*-score for x_i
    - $\bar{x}$ = sample mean
    - s = sample standard deviation

z-Scores for the Class Size Data

Number of Students in Class (x_i)	Deviation About the Mean ($x_i - \bar{x}$)	z-Score $\left(\frac{x_i - \bar{x}}{s}\right)$
46	2	$2/8 = .25$
54	10	$10/8 = 1.25$
42	−2	$-2/8 = -.25$
46	2	$2/8 = .25$
32	−12	$-12/8 = -1.50$

- For class size data, $\bar{x} = 44$ and $s = 8$.
  - For observations with a value > mean, z-score > 0.
  - For observations with a value < mean, z-score < 0.

Calculating *z*-Scores for the Home Sales Data in Excel

	A	B	C
1	Home Sale	Selling Price ($)	*z*-Score
2	1	138000	=STANDARDIZE(B2,B15,B16)
3	2	254000	=STANDARDIZE(B3,B15,B16)
4	3	186000	=STANDARDIZE(B4,B15,B16)
5	4	257500	=STANDARDIZE(B5,B15,B16)
6	5	108000	=STANDARDIZE(B6,B15,B16)
7	6	254000	=STANDARDIZE(B7,B15,B16)
8	7	138000	=STANDARDIZE(B8,B15,B16)
9	8	298000	=STANDARDIZE(B9,B15,B16)
10	9	199500	=STANDARDIZE(B10,B15,B16)
11	10	208000	=STANDARDIZE(B11,B15,B16)
12	11	142000	=STANDARDIZE(B12,B15,B16)
13	12	456250	=STANDARDIZE(B13,B15,B16)
14			
15	Mean:	=AVERAGE(B2:B13)	
16	Standard Deviation:	=STDEV.S(B2:B13)	

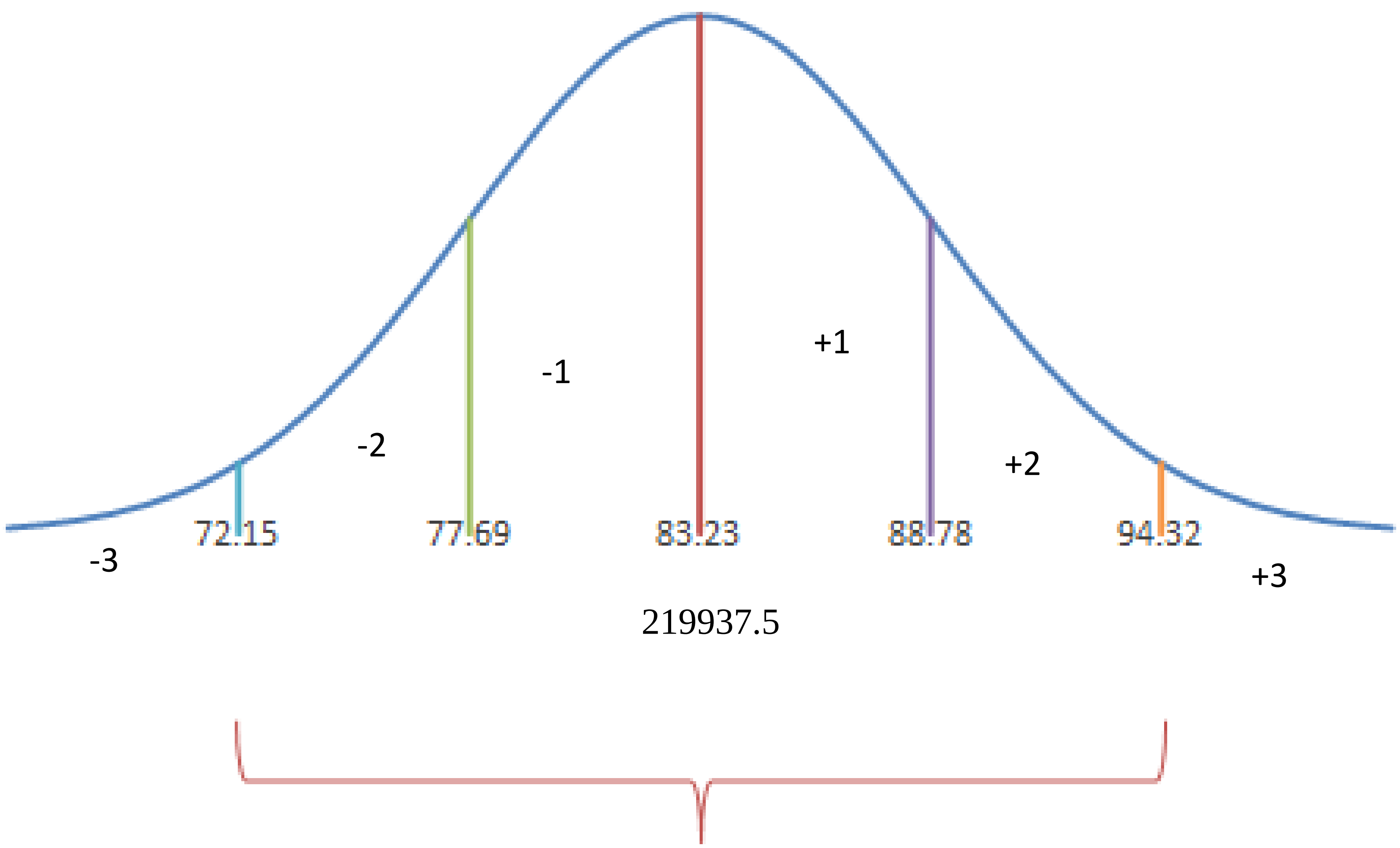
	A	B	C
1	Home Sale	Selling Price ($)	*z*-Score
2	1	138,000	–0.862
3	2	254,000	0.358
4	3	186,000	–0.357
5	4	257,500	0.395
6	5	108,000	–1.177
7	6	254,000	0.358
8	7	138,000	–0.862
9	8	298,000	0.821
10	9	199,500	–0.215
11	10	208,000	–0.126
12	11	142,000	–0.820
13	12	456,250	2.486
14			
15	Mean:	$ 219,937.50	
16	Standard Deviation:	$ 95,065.77	

EMPIRICAL RULE

- **Empirical rule**:
  - For data having a bell-shaped distribution:
    - Within 1 standard deviation – approximately 68% of the data values.
    - Within 2 standard deviations – approximately 95% of the data values.
    - Within 3 standard deviations – almost all the data values.
- Identifying outliers:
  - **Outliers**: Extreme values in a data set.
  - It can be identified using standardized values (*z*-scores).
    - Any data value with a *z*-score less than –3 or greater than +3 is an outlier.

COLLEGE OF CO

Bell Curve Chart



Prepared b