Chapter 3, Part A
Descriptive Statistics: Numerical Measures

- Measures of Location
- Measures of Variability

CENGAGE

Numerical Measures

- If the measures are computed for data from a sample, they are called sample statistics.
- If the measures are computed for data from a population, they are called population parameters.
- A sample statistic is referred to as the point estimator of the corresponding population parameter.

Measures of Location

- Mean
- Median
- Mode
- Weighted Mean
- Geometric Mean
- Percentiles
- Quartiles

Mean

Perhaps the most important measure of location is the mean.

- The mean provides a measure of central location.
- The mean of a data set is the average of all the data values.
- The sample mean $\bar{x}$ is the point estimator of the population mean μ.

Sample Mean $\bar{x}$

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

where:

$\sum x_i$ = sum of the values of the *n* observations

n = number of observations in the sample

Population Mean μ

$$\mu = \frac{\sum x_i}{n}$$

where:

$\sum x_i$ = sum of the values of the *n* observations

n = number of observations in the population

Sample Mean $\bar{x}$ (1 of 2)

Example: Monthly Starting Salary

A placement office wants to know the average starting salary of business graduates. Monthly starting salaries for a sample of 12 business school graduates is provided here.

Graduate	Monthly Starting Salary ($)	Graduate	Monthly Starting Salary ($)
1	5850	7	5890
2	5950	8	6130
3	6050	9	5940
4	5880	10	6325
5	5755	11	5920
6	5710	12	5880

Sample Mean $\bar{x}$ (2 of 2)

Example: Monthly Starting Salary

$$\bar{x} = \frac{\sum x_i}{n} = \frac{71{,}280}{12} = 5{,}940$$

Median (1 of 4)

- The median of a data set is the value in the middle when the data items are arranged in ascending order.
- Whenever a data set has extreme values, median is the preferred measure of central location.
- The median is the measure of location most often reported for annual income and property value data.
- A few extremely large incomes or property values can inflate the mean.

Median (2 of 4)

For an odd number of observations:

7 observations

26	18	27	12	14	27	19

12	14	18	19	26	27	27

In ascending order

Median is the middle value

Median = 19

CENGAGE

Median (3 of 4)

For an even number of observations:

8 observations

26	**18**	**27**	**12**	**14**	**27**	**19**	**30**

12	**14**	**18**	**19**	**26**	**27**	**27**	**30**

In ascending order

Median is the average of the middle two values.

$$\text{Median} = \frac{(19+26)}{2} = 22.5$$

Median (4 of 4)

<u>**Example:**</u> Monthly Starting Salary
Averaging the 6th and 7th
data values:

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

Note: The data is in ascending order.

$$\text{Median} = \frac{(5{,}890 + 5{,}920)}{2} = 5{,}905$$

Trimmed Mean

- Another measure sometimes used when extreme values are present is the trimmed mean.
- It is obtained by deleting a percentage of the smallest and largest values from a data set and then computing the mean of the remaining values.
- For example, the 5% trimmed mean is obtained by removing the smallest 5% and the largest 5% of the data values and then computing the mean of the remaining values.

Mode (1 of 2)

- The mode of a data set is the value that occurs with greatest frequency.
- The greatest frequency can occur at two or more different values.
- If the data have exactly two modes, the data are bimodal.
- If the data have more than two modes, the data are multimodal.

Mode (2 of 2)

Example: Monthly Starting Salary

The only monthly starting salary that occurs more than once is $5,880.

Mode = 5,880

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

Note: The data is in ascending order.

Using Excel to Compute Mean, Median, and Mode (1 of 2)

Excel's Mean Function

- AVERAGE(data cell range)

Excel's Median Function

- MEDIAN(data cell range)

Excel's Mode Function

- MODE.SNGL(data cell range)

Using Excel to Compute Mean, Median, and Mode (2 of 2)

- Excel Formula and Value Worksheets

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)			
2	1	5850		Mean	=AVERAGE(B2:B13)
3	2	5950		Median	=MEDIAN(B2:B13)
4	3	6050		Mode	=MODE.SNGL(B2:B13)
5	4	5880			
6	5	5755			
7	6	5710			
8	7	5890			
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)			
2	1	5850		Mean	5940
3	2	5950		Median	5905
4	3	6050		Mode	5880
5	4	5880			
6	5	5755			
7	6	5710			
8	7	5890			
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

Weighted Mean (1 of 4)

- In some instances the mean is computed by giving each observation a weight that reflects its relative importance.
- The choice of weights depends on the application.
- The weights might be the number of credit hours earned for each grade, as in GPA.
- In other weighted mean computations, quantities such as pounds, dollars, or volume are frequently used.

Weighted Mean (2 of 4)

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

where: x_i = value of observation *i*

w_i = weight for observation *i*

Numerator: sum of the weighted data values

Denominator: sum of the weights

If data is from a population, μ replaces $\bar{x}$.

Weighted Mean (3 of 4)

Example: Purchase of Raw Material

Consider the following sample of five purchases of a raw material over a period of three months:

Purchase	Cost per Pound ($)	Number of Pounds
1	3.00	1200
2	3.4	500
3	2.8	2750
4	2.9	1000
5	3.25	800

CENGAGE

Weighted Mean (4 of 4)

Example: Purchase of raw material

Purchase	Cost per Pound ($) x_i	Number of Pounds w_i	$w_i x_i$
1	3.00	1200	3600
2	3.4	500	1700
3	2.8	2750	7700
4	2.9	1000	2900
5	255755	800	2600

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{18{,}500}{6{,}250} = 2.96 = \$2.96$$

FYI, equally weighted (simple) mean = $3.07

Geometric Mean (1 of 4)

- The geometric mean is calculated by finding the *n*th root of the product of *n* values.
- It is often used in analyzing growth rates in financial data (where using the arithmetic mean will provide misleading results).
- It should be applied anytime you want to determine the mean rate of change over several successive periods (be it years, quarters, weeks, . . .).
- Other common applications include changes in populations of species, crop yields, pollution levels, and birth and death rates.

Geometric Mean (2 of 4)

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\ldots(x_n)}$$

$$= \left[(x_1)(x_2)\ldots(x_n)\right]^{1/n}$$

Geometric Mean (3 of 4)

Example: Mutual fund

Year	Annual Return %	Growth Factor
1	−22.1	0.779
2	28.7	1.287
3	10.9	1.109
4	4.9	1.049
5	15.8	1.158
6	5.5	1.055
7	−37.0	0.630
8	26.5	1.265
9	15.1	1.151
10	2.1	1.021

CENGAGE

Geometric Mean (4 of 4)

$$\bar{x}_g = \sqrt[10]{(0.779)(1.287)(1.109)(1.049)(1.158)(1.055)(0.630)(1.265)(1.151)(1.021)}$$

$$= \sqrt[10]{1.334493}$$

$$= 1.029275$$

Average growth rate per period is $(1.029275 - 1)(100) = 2.9\%$

Using Excel to Compute Geometric Mean (1 of 2)

- Excel Function for Geometric Mean

= GEOMEAN(data cell range)

Using Excel to Compute Geometric Mean (2 of 2)

- Excel Formula and Value Worksheets

	A	B	C	D	E	F	G
1	Year	Return (%)	Growth Factor				
2	1	-22.1	=1+0.01*B2		Geometric Mean	=GEOMEAN(C2:C11)	
3	2	28.7	=1+0.01*B3				
4	3	10.9	=1+0.01*B4				
5	4	4.9	=1+0.01*B5				
6	5	15.8	=1+0.01*B6				
7	6	5.5	=1+0.01*B7				
8	7	-37	=1+0.01*B8				
9	8	26.5	=1+0.01*B9				
10	9	15.1	=1+0.01*B10				
11	10	2.1	=1+0.01*B11				
12							

	A	B	C	D	E	F	G
1	Year	Return (%)	Growth Factor				
2	1	-22.1	0.779		Geometric Mean	1.029275	
3	2	28.7	1.287				
4	3	10.9	1.109				
5	4	4.9	1.049				
6	5	15.8	1.158				
7	6	5.5	1.055				
8	7	-37	0.63				
9	8	26.5	1.265				
10	9	15.1	1.151				
11	10	2.1	1.021				
12							

CENGAGE

Percentiles (1 of 2)

- A percentile provides information about how the data are spread over the interval from the smallest value to the largest value.
- Admission test scores for colleges and universities are frequently reported in terms of percentiles.
- The *p*th percentile of a data set is a value such that at least p% of the items take on this value or less and at least $(100 - p)\%$ of the items take on this value or more.

Percentiles (2 of 2)

Arrange the data in ascending order.

Compute L_p, the location of the *p*th percentile.

$$L_p = \left(\frac{p}{100}\right)(n+1)$$

80th Percentile (1 of 2)

Example: Monthly Starting Salary

$$L_p = (p/100)(n+1) = (80/100)(12+1) = 10.4$$

(the 10th value plus .4 times the difference between the 11th and 10th values)

$$\text{80th Percentile} = 6{,}050 + 0.4(6{,}130 - 6{,}050) = 6{,}082$$

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

CENGAGE

80th Percentile (2 of 2)

Example: Monthly Starting Salary

At least 80% of the items take on a value of 6082 or less.
10/12 = .833 or 83%

At least 20% of the items take on a value of 6082 or more.
2/12 = .167 or 16.7%

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

CENGAGE

Quartiles

Quartiles are specific percentiles.

- First Quartile = 25th Percentile
- Second Quartile = 50th Percentile = Median
- Third Quartile = 75th Percentile

Third Quartile (75th Percentile)

<u>Example:</u> Monthly Starting Salary

$$L_p = (p/100)(n+1) = (75/100)(12+1) = 9.75$$

(the 9th value plus .75 times the difference between the 10th and 9th values)

$$\text{Third quartile} = 5{,}950 + .75(6{,}050 - 5{,}950) = 6{,}025$$

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

Using Excel to Compute Percentiles and Quartiles

- Excel's Percentile Function

    PERCENTILE.EXC(data range, *p*/100)

- Excel's Quartile Function

    QUARTILE.EXC(array, QUART)

Using Excel to Compute 80th Percentile and Quartiles

- Excel Formula and Value Worksheets

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)		Percentile	Value
2	1	5850		80	=PERCENTILE.EXC(B2:B13,D2/100)
3	2	5950			
4	3	6050			
5	4	5880		Quartile	Value
6	5	5755		1	=QUARTILE.EXC($B$2:$B$13,D6)
7	6	5710		2	=QUARTILE.EXC($B$2:$B$13,D7)
8	7	5890		3	=QUARTILE.EXC($B$2:$B$13,D8)
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)		Percentile	Value
2	1	5850		80	6082.00
3	2	5950			
4	3	6050			
5	4	5880		Quartile	Value
6	5	5755		1	5857.50
7	6	5710		2	5905.00
8	7	5890		3	6025.00
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

Note: It is not necessary to put the data in ascending order.

Measures of Variability (1 of 2)

- It is often desirable to consider measures of variability (dispersion), as well as measures of location.
- For example, in choosing supplier A or supplier B we might consider not only the average delivery time for each but also the variability in delivery time for each.

Measures of Variability (2 of 2)

- Range
- Interquartile Range
- Variance
- Standard Deviation
- Coefficient of Variation

Range (1 of 2)

- The range of a data set is the difference between the largest and smallest data values.

$$\text{Range} = \text{Largest value} - \text{Smallest value}$$

- It is the simplest measure of variability.
- It is very sensitive to the smallest and largest data values.

Range (2 of 2)

Example: Monthly Starting Salary

$$\text{Range} = \text{largest value} - \text{smallest value}$$

$$\text{Range} = 6{,}325 - 5{,}710 = 615$$

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

Interquartile Range

- The interquartile range of a data set is the difference between the third quartile and the first quartile.
- It is the range for the middle 50% of the data.
- It overcomes the sensitivity to extreme data values.

Interquartile Range (IQR)

Example: Monthly Starting Salary

- 3rd Quartile $(Q_3) = 6{,}000$
- 1st Quartile $(Q_1) = 5{,}865$

- $\text{IQR} = Q_3 - Q_1 = 6{,}000 - 5{,}865 = 135$

5,710	5,755
5,850	5,880
5,880	5,890
5,920	5,940
5,950	6,050
6,130	6,325

CENGAGE

Variance (1 of 2)

- The variance is a measure of variability that utilizes all the data.
- It is based on the difference between the value of each observation (x_i) and the mean ($\bar{x}$ for a sample, μ for a population).
- The variance is useful in comparing the variability of two or more variables.

Variance (2 of 2)

- The variance is the average of the squared differences between each data value and the mean.
- The variance is computed as follows:

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

for a sample

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

for a population

Standard Deviation (1 of 2)

- The standard deviation of a data set is the positive square root of the variance.
- It is measured in the same units as the data, making it more easily interpreted than the variance.

Standard Deviation (2 of 2)

The standard deviation is computed as follows:

For a sample	For a population
$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$

Excel's Variance and Standard deviation Functions

- Excel's Sample Variance Function

  VAR.S(data cell range)

- Excel's Sample Standard deviation Function

  STDEV.S(data cell range)

CENGAGE

Coefficient of Variation

- The coefficient of variation indicates how large the standard deviation is in relation to the mean.

The coefficient of variation is computed as follows:

$$\left[\frac{s}{\bar{x}} \times 100\right]\%$$

for a sample

$$\left[\frac{\sigma}{\mu} \times 100\right]\%$$

for a population

CENGAGE

Sample Variance, Standard Deviation, And Coefficient of Variation

Example: Monthly Starting Salary

Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 27,440.91$$

Standard Deviation

$$s = \sqrt{s^2} = \sqrt{27,440.91} = 165.65$$

Coefficient of Variation

$$\left[\frac{s}{\bar{x}} \times 100\right]\% = \left[\frac{165.65}{3,940} \times 100\right]\% = 4.2\%$$

Using Excel to compute the sample Variance, Standard Deviation and Coefficient of Variation

- Excel Formula and Value Worksheets

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)			
2	1	5850		Mean	=AVERAGE(B2:B13)
3	2	5950		Median	=MEDIAN(B2:B13)
4	3	6050		Mode	=MODE.SNGL(B2:B13)
5	4	5880		Variance	=VAR.S(B2:B13)
6	5	5755		Standard Deviation	=STDEV.S(B2:B13)
7	6	5710			
8	7	5890			
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

	A	B	C	D	E
1	Graduate	Monthly Starting Salary ($)			
2	1	5850		Mean	5940
3	2	5950		Median	5905
4	3	6050		Mode	5880
5	4	5880		Variance	27440.91
6	5	5755		Standard Deviation	165.65
7	6	5710			
8	7	5890			
9	8	6130			
10	9	5940			
11	10	6325			
12	11	5920			
13	12	5880			
14					

CENGAGE