

Describing Data: Numerical Measures



Chapter 3

GOALS

1. Calculate the *arithmetic mean, weighted mean, median, mode, and geometric mean*.
2. Explain the characteristics, uses, advantages, and disadvantages of each *measure of location*.
3. Compute and understand *quartiles, deciles, and percentiles*.

Numerical Descriptive Measures

Measures of Location

- Arithmetic Mean
- Weighted Mean
- Median
- Mode
- Geometric Mean

Measures of Dispersion

- Range
- Mean Deviation
- Variance
- Standard Deviation

Population Mean

For ungrouped data, the **population mean** is the sum of all the population values divided by the total number of population values:

POPULATION MEAN

$$\mu = \frac{\sum X}{N}$$

[3-1]

where:

- μ represents the population mean. It is the Greek lowercase letter “mu.”
- N is the number of values in the population.
- X represents any particular value.
- Σ is the Greek capital letter “sigma” and indicates the operation of adding.
- ΣX is the sum of the X values in the population.

EXAMPLE – Population Mean

There are 12 automobile manufacturing companies in the United States. Listed below is the number of patents granted by the United States government to each company in a recent year.

Company	Number of Patents Granted	Company	Number of Patents Granted
General Motors	511	Mazda	210
Nissan	385	Chrysler	97
DaimlerChrysler	275	Porsche	50
Toyota	257	Mitsubishi	36
Honda	249	Volvo	23
Ford	234	BMW	13

Is this information a sample or a population? What is the arithmetic mean number of patents granted?

$$\mu = \frac{\sum X}{N} = \frac{511 + 385 + 275 + \dots + 36 + 23 + 13}{12} = \frac{2340}{12} = 195$$

Sample Mean

- For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values:

SAMPLE MEAN

$$\bar{X} = \frac{\sum X}{n}$$

[3-2]

where:

$\bar{X}$ is the sample mean. It is read “X bar.”

n is the number of values in the sample.

EXAMPLE – Sample Mean

SunCom is studying the number of minutes used monthly by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

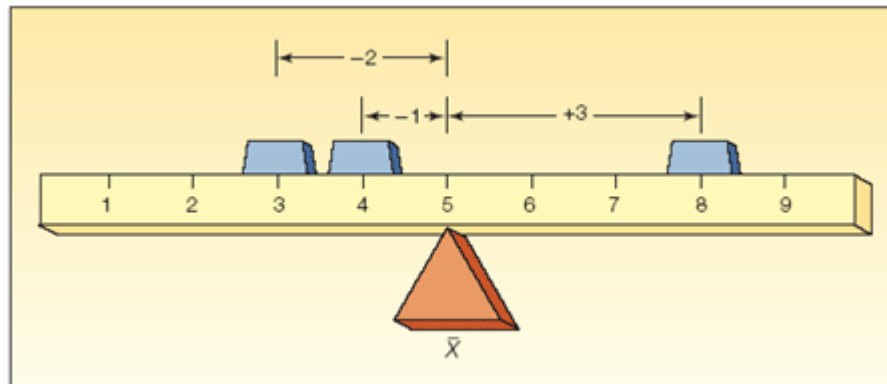
90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

$$\bar{X} = \frac{\sum X}{n} = \frac{90 + 77 + \dots + 83}{12} = \frac{1170}{12} = 97.5$$

Properties of the Arithmetic Mean

1. Every set of interval-level and ratio-level data has a mean.
2. All the values are included in computing the mean.
3. The mean is unique.
4. The sum of the deviations of each value from the mean is zero.



Weighted Mean

- The **weighted mean** of a set of numbers $X_1, X_2, \dots, X_n$, with corresponding weights $w_1, w_2, \dots, w_n$, is computed from the following formula:

WEIGHTED MEAN

$$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \dots + w_nX_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

[3-3]

EXAMPLE – Weighted Mean

The Carter Construction Company pays its hourly employees \$16.50, \$19.00, or \$25.00 per hour. There are 26 hourly employees, 14 of which are paid at the \$16.50 rate, 10 at the \$19.00 rate, and 2 at the \$25.00 rate. What is the mean hourly rate paid the 26 employees?

$$\bar{X}_w = \frac{14(\$16.50) + 10(\$19.00) + 2(\$25.00)}{14 + 10 + 2} = \frac{\$471.00}{26} = \$18.1154$$

The Median

The **Median** is the midpoint of the values after they have been ordered from the smallest to the largest.

- There are as many values above the median as below it in the data array.
- For an even set of values, the median will be the arithmetic average of the two middle numbers.

Properties of the Median

1. There is a unique median for each data set.
2. It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.
3. It can be computed for ratio-level, interval-level, and ordinal-level data.
4. It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended class.

EXAMPLES - Median

The ages for a sample of five college students are:

21, 25, 19, 20, 22

Arranging the data in ascending order gives:

19, 20, 21, 22, 25.

Thus the median is 21.

The heights of four basketball players, in inches, are:

76, 73, 80, 75

Arranging the data in ascending order gives:

73, 75, 76, 80.

Thus the median is 75.5

The Mode

- The **mode** is the value of the observation that appears most frequently.

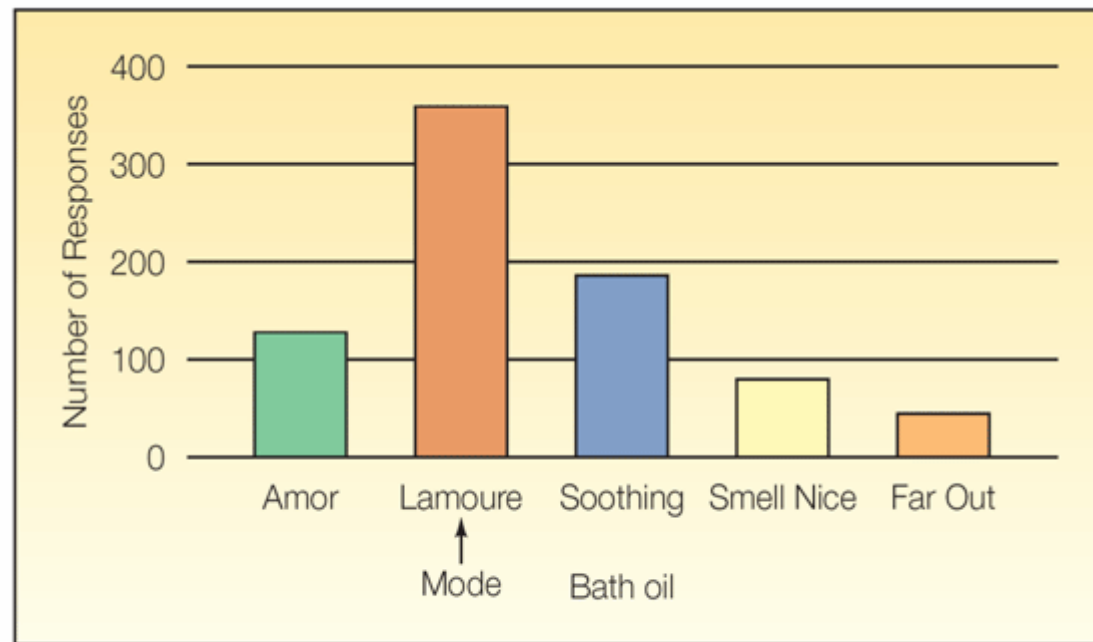


CHART 3-1 Number of Respondents Favoring Various Bath Oils

Example - Mode

The annual salaries of quality-control managers in selected states are shown below. What is the modal annual salary?

State	Salary	State	Salary	State	Salary
Arizona	\$35,000	Illinois	\$58,000	Ohio	\$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

A perusal of the salaries reveals that the annual salary of \$60,000 appears more often (six times) than any other salary. The mode is, therefore, \$60,000.

Mean, Median, Mode Using Excel

Table 2–4 in Chapter 2 shows the prices of the 80 vehicles sold last month at Whitner Autoplex in Raytown, Missouri. Determine the mean and the median selling price.

The mean and the median selling prices are reported in the following Excel output. There are 80 vehicles in the study. So the calculations with a calculator would be tedious and prone to error.



TABLE 2–4 Prices of Vehicles Sold Last Month at Whitner Autoplex

\$23,197	\$23,372	\$20,454	\$23,591	\$26,651	\$27,453	\$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Lowest

Highest

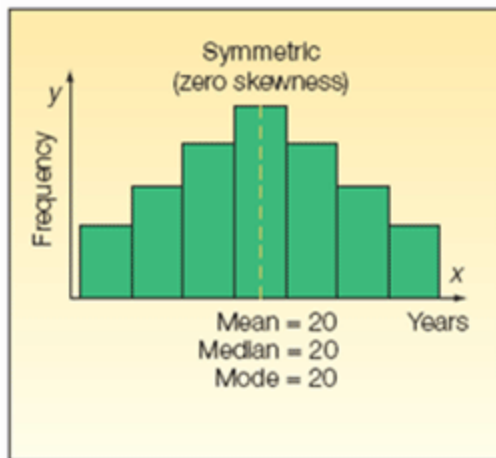
Mean, Median, Mode Using Excel

The screenshot shows a Microsoft Excel window titled "Table2-1 [Read-Only]". The data is organized in columns A through D, with headers "Price", "Price(\$000)", "Age", and "Type". The data rows are numbered 1 through 20. A summary statistics box is overlaid on the right side of the spreadsheet, titled "Price", and lists various statistical measures for the "Price" column. The values for Mean, Median, and Mode are highlighted in yellow. The status bar at the bottom indicates the current cell is E8, and the sheet is "Sheet3".

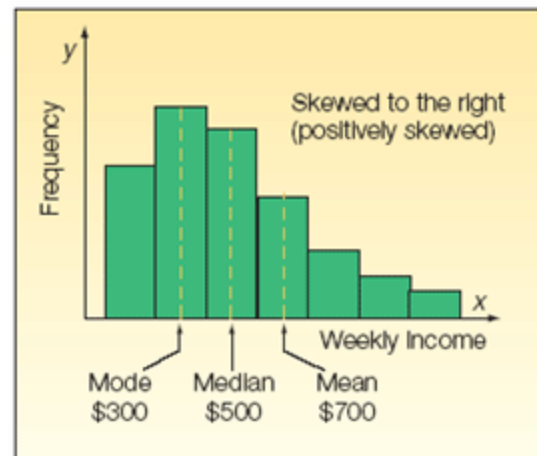
	A	B	C	D
1	Price	Price(\$000)	Age	Type
2	23197	23.197	48	0
3	23372	23.372	48	0
4	20454	20.454	40	1
5	23591	23.591	40	0
6	26851	26.851	46	1
7	27453	27.453	37	1
8	17266	17.266	32	1
9	18021	18.021	29	1
10	28683	28.683	38	1
11	30872	30.872	43	0
12	18587	18.587	32	0
13	23168	23.168	47	0
14	35851	35.851	56	0
15	19251	19.251	42	1
16	20047	20.047	28	1
17	24205	24.205	56	0
18	24324	24.324	50	1
19	24609	24.609	31	1
20	28670	28.67	51	1

Price	
Mean	23218.1625
Standard Error	486.9409474
Median	22831
Mode	20642
Standard Deviation	4354.43781
Sample Variance	18961128.64
Kurtosis	0.5433087
Skewness	0.72681585
Range	20378
Minimum	15546
Maximum	35825
Sum	1857453
Count	80

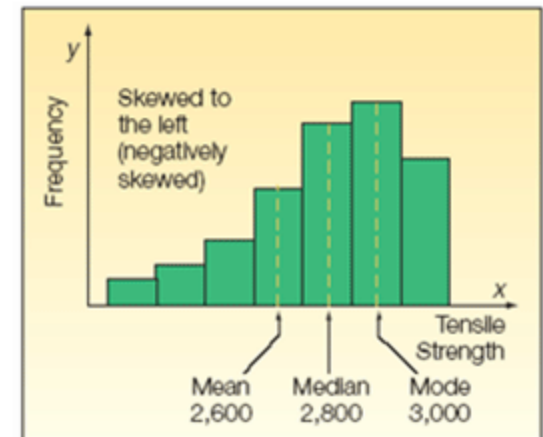
The Relative Positions of the Mean, Median and the Mode



zero skewness
mode = median = mean



positive skewness
mode < median < mean



negative skewness
mode > median > mean

The Geometric Mean

- Useful in finding the average change of percentages, ratios, indexes, or growth rates over time.
- Has a wide application in business and economics because we are often interested in finding the percentage changes in sales, salaries, or economic figures, such as the GDP, which compound or build on each other.
- Will always be less than or equal to the arithmetic mean.
- Defined as the n th root of the product of n values.
- The formula for the geometric mean is written:

GEOMETRIC MEAN

$$GM = \sqrt[n]{(X_1)(X_2) \cdots (X_n)}$$

[3–4]

EXAMPLE – Geometric Mean

The return on investment earned by Atkins construction Company for four successive years was: 30 percent, 20 percent, 40 percent, and 200 percent. What is the geometric mean rate of return on investment?

$$GM = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

Percentiles - Example

Listed below are the commissions earned last month by a sample of 15 brokers at Salomon Smith Barney's Oakland, California, office. Salomon Smith Barney is an investment company with offices located throughout the United States.

\$2,038	\$1,758	\$1,721	\$1,637
\$2,097	\$2,047	\$2,205	\$1,787
\$2,287	\$1,940	\$2,311	\$2,054
\$2,406	\$1,471	\$1,460	

Locate the **median**, the **first quartile**, and the **third quartile** for the commissions earned.

Percentiles – Example (cont.)

Step 1: Organize the data from lowest to largest value

\$1,460	\$1,471	\$1,637	\$1,721
\$1,758	\$1,787	\$1,940	\$2,038
\$2,047	\$2,054	\$2,097	\$2,205
\$2,287	\$2,311	\$2,406	

Other Measures of Dispersion: Quartiles, Deciles and Percentiles

- The standard deviation is the most widely used measure of dispersion.
- Alternative ways of describing spread of data include determining the *location* of values that divide a set of observations into equal parts.

LOCATION OF A PERCENTILE

$$L_p = (n + 1) \frac{P}{100}$$

[4-1]

- These measures include **quartiles, deciles, and percentiles.**

Percentile Computation

- Let L_p refer to the location of a desired percentile. If we wanted to find the 33rd percentile we would use L_{33} and if we wanted the median, the 50th percentile, then L_{50} .

LOCATION OF A PERCENTILE

$$L_p = (n + 1) \frac{P}{100}$$

[4-1]

- The number of observations is n . To locate the median, its position is at $(n + 1)/2$. We could write this as $(n + 1)(P/100)$, where P is the desired percentile.

Percentiles – Example (cont.)

Step 2: Compute the first and third quartiles. Locate L_{25} and L_{75} using:

LOCATION OF A PERCENTILE

$$L_p = (n + 1) \frac{P}{100}$$

[4-1]

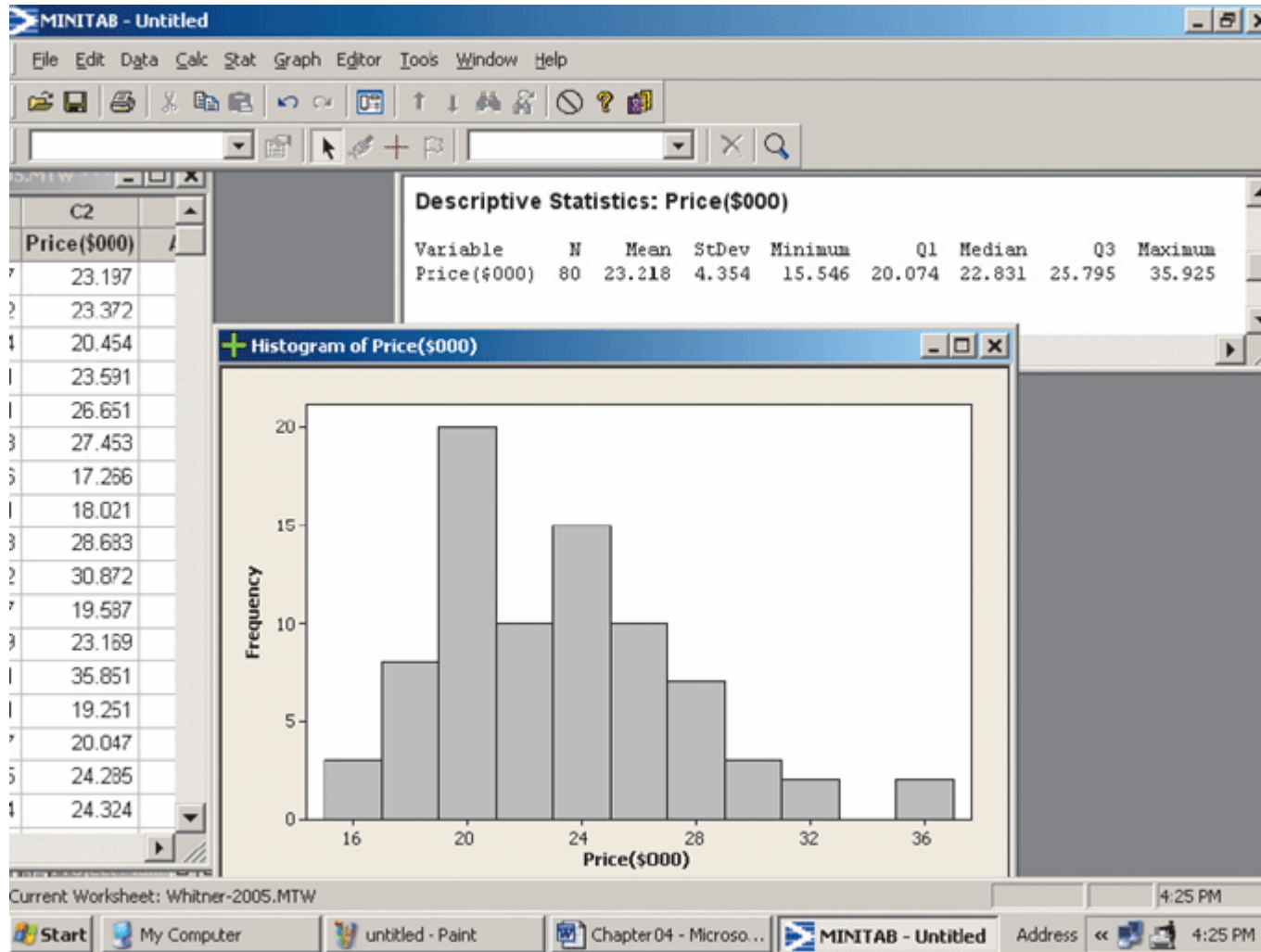
$$L_{25} = (15 + 1) \frac{25}{100} = 4 \qquad L_{75} = (15 + 1) \frac{75}{100} = 12$$

Therefore, the first and third quartiles are the 4th and 12th observations in the array, respectively

$$L_{25} = \$1,721$$

$$L_{75} = \$2,205$$

Percentiles – Example (Minitab)



Percentiles – Example (Excel)

Microsoft Excel - WHITNER-2005

File Edit View Insert Format Tools MegaStat Data Window Help

MS Sans Serif 10

MS Sans Serif 10 B I U

J24

	B	C	D	E	F	G	H	I	J
1	Price(\$000)	Age	Type				Price(\$000)		
2	23.197	46	0						
3	23.372	48	0				Mean	23.2181625	
4	20.454	40	1				Standard Error	0.486840947	
5	23.591	40	0				Median	22.831	
6	26.651	46	1				Mode	20.642	
7	27.453	37	1				Standard Deviation	4.35443781	
8	17.266	32	1				Sample Variance	18.96112864	
9	18.021	29	1				Kurtosis	0.5433087	
10	28.683	38	1				Skewness	0.72681585	
11	30.872	43	0				Range	20.379	
12	19.587	32	0				Minimum	15.546	
13	23.169	47	0				Maximum	35.925	
14	35.851	56	0				Sum	1857.453	
15	19.251	42	1				Count	80	
16	20.047	28	1				Largest(20)	25.799	
17	24.285	56	0				Smallest(20)	20.047	
18	24.324	50	1						
19	24.609	31	1						
20	28.67	51	1						

Output Sheet1

Ready NUM

Start 2 Microsoft... Chapter04 Chapter04x shot4-4 - Paint shot4-3 - Win... Address 10:55 AM