# Chapter One

**1.1**    The word 'statistics' has the following two meanings.

   i. First, it refers to numerical facts such as the ages of persons, incomes of families, etc.

   ii. Second, it refers to the field of study. It provides us with techniques that help us to collect, analyze, present, and interpret data and to make decisions.

**1.2**    There are two types of statistics: descriptive statistics and inferential statistics. Descriptive statistics consists of methods that help us organize, display, and describe data using tables, graphs, and summary measures. Inferential statistics is the part of statistics that helps us make decisions about the population based on sample results.

**1.3**    **Population:** All elements whose characteristics are being studied.

   **Sample:** A portion of the population selected for a study.

   **Representative sample:** A sample that possesses the characteristics of the population as closely as possible.

   **Random sample:** A sample drawn in such a way that each element of the population has some chance of being included in the sample.

   **Sampling with replacement:** A sampling procedure in which the item selected at any selection is put back in the population before the next item is drawn.

   **Sampling without replacement:** A sampling procedure in which the item selected at each selection is not replaced in the population.

**1.4**    i. Suppose the employees of a company select one employee from all employees to be on the management board. Then, they select one employee from the same group of all employees to be a representative on the workers union. This is an example of sampling with replacement.

   ii. Suppose the employees of a company select one employee from all employees to be on the management board. Then, they exclude the selected employee from the list of all employees and select one employee from the remaining employees to be a representative on the workers union. This is an example of sampling without replacement.

**1.5**    A census is a survey that includes all members of the population. A survey based on a portion of the population is called a sample survey. A sample survey is preferred over a census for the following reasons.

   i. Conducting a census is very expensive because the size of the population is usually very large.

   ii. Conducting a census is very time consuming.

iii. In many cases it is almost impossible to identify every member of the target population.

**1.6**  a. Population  b. Population  c. Sample  d. Population  e. Sample

**1.7**  a. Population  b. Sample  c. Population  d. Sample  e. Population

**1.8**  **Element:** A specific subject or object on which the information is collected.

**Variable:** A characteristic that is being investigated and that can assume different values for different elements.

**Observation:** The value of a variable for a single element.

**Data set:** A collection of observations on one or more variables.

**1.9**  **Member:** Each student listed in the table.

**Variable:** Score on the statistics test.

**Measurement:** Score of a specific student. For example, Susan's score of 91 is a measurement.

**Data set:** Collection of scores for the five students listed in the table.

**1.10**  **Member:** Each state included in the table.

**Variable:** Amount of garbage (in millions of tons).

**Measurement:** Amount of garbage for each state. For example, the amount of garbage for Texas is 33.8 million tons, which is a measurement.

**Data set:** Collection of amounts of garbage for the eight states listed in the table.

**1.11**  a. Scores  b. Five observations  c. Five elements (students)

**1.12**  a. Amount of garbage  b. Eight  c. Eight elements (states)

**1.13**  a.  **Quantitative variable**: A variable that can assume numerical values.

b.  **Qualitative variable**: A variable that cannot be measured numerically but can be divided into different categories.

c.  **Discrete variable**: A variable whose values are countable.

d.  **Continuous variable**: A variable that can assume any value over a certain interval or intervals.

e.  **Quantitative data**: Data collected on a quantitative variable.

f.  **Qualitative data**: Data collected on a qualitative variable.

**1.14**  a. Quantitative  b. Qualitative  c. Qualitative  d. Quantitative  e. Quantitative

**1.15**  a. Quantitative  b. Quantitative  c. Qualitative  d. Quantitative  e. Qualitative

**1.16**  a. Discrete  d. Continuous  e. Discrete

**1.17**  a. Discrete  b. Continuous  d. Continuous

**1.18**    Data collected on different elements for the same period of time are called cross-section data. Prices for twenty cars of 1997 models is an example of cross-section data.

Data collected on the same element for the same variable for different periods of time are called time-series data. Average prices of houses in Connecticut for the years 1970 to 1997 is an example of time-series data.

**1.19**    Internal sources of data are a company's own files and records.

External sources of data are the sources that do not belong to a company.

**1.20**    a. Time-series data        b. Cross-section data        c. Time-series data        d. Cross-section data

**1.21**    a. Cross-section data        b. Cross-section data        c. Time-series data        d. Time-series data

**1.22**

| $m$ | $f$ | $f^2$ | $mf$ | $m^2f$ |
|---|---|---|---|---|
| 5 | 12 | 144 | 60 | 300 |
| 10 | 8 | 64 | 80 | 800 |
| 7 | 6 | 36 | 102 | 1734 |
| 20 | 16 | 256 | 320 | 6400 |
| 25 | 4 | 16 | 100 | 2500 |
| Sum=77 | 46 | 516 | 662 | 11,734 |

a. $\sum m = 77$        b. $\sum f^2 = 516$        c. $\sum mf = 662$        d. $\sum m^2f = 11{,}734$

**1.23**

| $m$ | $f$ | $m^2$ | $mf$ | $m^2f$ |
|---|---|---|---|---|
| 3 | 16 | 9 | 48 | 144 |
| 6 | 11 | 36 | 66 | 396 |
| 25 | 16 | 625 | 400 | 10,000 |
| 12 | 8 | 144 | 96 | 1152 |
| 15 | 4 | 225 | 60 | 900 |
| 18 | 14 | 324 | 252 | 4536 |
| Sum= 79 | 69 | 1363 | 922 | 17,128 |

a. $\sum f = 69$        b. $\sum m^2 = 1363$        c. $\sum mf = 922$        d. $\sum m^2f = 17{,}128$

**1.24**

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 15 | 10 | 150 | 225 | 100 |
| 22 | 12 | 264 | 484 | 144 |
| 11 | 14 | 154 | 121 | 196 |
| 8 | 9 | 72 | 64 | 81 |
| 5 | 18 | 90 | 25 | 324 |
| $\sum x = 61$ | $\sum y = 63$ | $\sum xy = 730$ | $\sum x^2 = 919$ | $\sum y^2 = 845$ |

a. $\sum x = 61$        b. $\sum y = 63$        c. $\sum xy = 730$        d. $\sum x^2 = 919$        e. $\sum y^2 = 845$

**1.25**

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 4 | 12 | 48 | 16 | 144 |
| 18 | 5 | 90 | 324 | 25 |
| 25 | 14 | 350 | 625 | 196 |
| 9 | 7 | 63 | 81 | 49 |
| 12 | 12 | 144 | 144 | 144 |
| 20 | 8 | 160 | 400 | 64 |
| $\sum x = 88$ | $\sum y = 58$ | $\sum xy = 855$ | $\sum x^2 = 1590$ | $\sum y^2 = 622$ |

a. $\sum x = 88$      b. $\sum y = 58$      c. $\sum xy = 855$      d. $\sum x^2 = 1590$      e. $\sum y^2 = 622$

**1.26**    a.   $\sum x = 75 + 80 + 97 + 91 + 63 = 406$

       b.   $(\sum x)^2 = (406)^2 = 164{,}836$

       c.   $\sum x^2 = (75)^2 + (80)^2 + (97)^2 + (91)^2 + (63)^2 = 33{,}684$

**1.27**    a.   $\sum x = 3 + 2 + 1 + 4 + 1 + 2 = 13$

       b.   $(\sum x)^2 = (13)^2 = 169$

       c.   $\sum x^2 = (3)^2 + (2)^2 + (1)^2 + (4)^2 + (1)^2 + (2)^2 = 35$

**1.28**    a.   $\sum x = 122 + 72 + 96 + 110 = \$400$

       b.   $(\sum x)^2 = (400)^2 = 160{,}000$

       c.   $\sum x^2 = (122)^2 + (72)^2 + (96)^2 + (110)^2 = 41{,}384$

**1.29**    a.   $\sum x = 7 + 9 + 6 + 12 + 10 + 9 + 8 = 61$ pounds

       b.   $(\sum x)^2 = (61)^2 = 3721$

       c.   $\sum x^2 = (7)^2 + (9)^2 + (6)^2 + (12)^2 + (10)^2 + (9)^2 + (8)^2 = 555$

**1.30**    **Variable:** Average math SAT scores.

       **Measurement:** Average math SAT score for a specific year.

       **Data Set:** Collection of average math SAT scores for six years listed in the table.

**1.31**    **Variable:** The governor's annual salary as of October 1998.

       **Measurement:** The governor's salary for a specific state.

       **Data Set:** Collection of the governors' salaries for the six states given in the table.

**1.32**    The data set in Exercise 1.30 is on the same variable (average math SAT score) for different periods of time. Hence, it is a time-series data set. The data set in exercise 1.31 is on the same variable (the governor's salary) for different states, but for the same period of time. Hence, it is a cross-section data set.

**1.33**    a. Sample        b. Population for that year     c. Sample        d. Population

**1.34**    a. Population        b. Sample            c. Sample        d. Population

**1.35**  a.  Sampling without replacement because once a patient is selected, he/she will not be replaced before the next patient is selected. All 10 selected patients must be different.

b.  Sampling with replacement because both times the selection is made from the same group (consisting of all professors).

**1.36**  a.  Sampling without replacement because once a student is selected he/she will not be replaced before the next student is selected. All five students selected must be different.

b.  Sampling with replacement because both times the selection is made from the same group (consisting of all students).

**1.37**  a.  $\sum x = 10 + 9 + 14 + 12 + 7 + 4 = 56$

b.  $(\sum x)^2 = (56)^2 = 3136$

c.  $\sum x^2 = (10)^2 + (9)^2 + (14)^2 + (12)^2 + (7)^2 + (4)^2 = 586$

**1.38**  a.  $\sum x = 43 + 39 + 44 + 31 + 40 = 197$

b.  $(\sum x)^2 = (197)^2 = 38{,}809$

c.  $\sum x^2 = (43)^2 + (39)^2 + (44)^2 + (31)^2 + (40)^2 = 7867$

**1.39**

| $m$ | $f$ | $f^2$ | $mf$ | $m^2 f$ | $m^2$ |
|---|---|---|---|---|---|
| 3 | 7 | 49 | 21 | 63 | 9 |
| 16 | 32 | 1024 | 512 | 8192 | 256 |
| 11 | 17 | 289 | 187 | 2057 | 121 |
| 9 | 12 | 144 | 108 | 972 | 81 |
| 20 | 34 | 1156 | 680 | 13,600 | 400 |
| Sum=59 | 102 | 2662 | 1508 | 24,884 | 867 |

a. $\sum m = 59$     b. $\sum f^2 = 2662$     c. $\sum mf = 1508$
d. $\sum m^2 f = 24{,}884$     e. $\sum m^2 = 867$

**1.40**

| $x$ | $y$ | $x^2$ | $xy$ | $x^2 y$ | $y^2$ |
|---|---|---|---|---|---|
| 7 | 5 | 49 | 35 | 245 | 25 |
| 11 | 15 | 121 | 165 | 1815 | 225 |
| 8 | 7 | 64 | 56 | 448 | 49 |
| 4 | 10 | 16 | 40 | 160 | 100 |
| 14 | 9 | 196 | 126 | 1764 | 81 |
| 28 | 19 | 784 | 532 | 14,896 | 361 |
| Sum=72 | 65 | 1230 | 954 | 19,328 | 841 |

a. $\sum y = 65$     b. $\sum x^2 = 1230$     c. $\sum xy = 954$
d. $\sum x^2 y = 19{,}328$     e. $\sum y^2 = 841$

## Self-Review Test for Chapter One

**1.** b

**2.** c

**3.** a. A sample without replacement

b. A sample with replacement

**4.** a. Qualitative

b. Quantitative; discrete

c. Quantitative; continuous

**5.** **Member:** A specific graduate included in the table. For example, Matt is a member.

**Variable:** Starting salaries of college graduates.

**Measurement:** Starting salary of a specific graduate. For example, Matt's starting salary of \$29,200 is a measurement.

**Data set:** The collection of starting salaries of five graduates listed in the table.

**6.** a.  $\sum x = 13 + 9 + 3 + 28 + 7 + 46 = 106$

b.  $\sum x^2 = (13)^2 + (9)^2 + (3)^2 + (28)^2 + (7)^2 + (46)^2 = 3208$

c.  $(\sum x)^2 = (106)^2 = 11{,}236$

**7.**

| $m$ | $f$ | $m^2$ | $mf$ | $m^2f$ | $f^2$ |
|---|---|---|---|---|---|
| 3 | 15 | 9 | 45 | 135 | 225 |
| 6 | 25 | 36 | 150 | 900 | 625 |
| 9 | 40 | 81 | 360 | 3240 | 1600 |
| 12 | 20 | 144 | 240 | 2880 | 400 |
| 15 | 12 | 225 | 180 | 2700 | 144 |
| $\sum m = 45$ | $\sum f = 112$ | $\sum m^2 = 495$ | $\sum mf = 975$ | $\sum m^2f = 9855$ | $\sum f^2 = 2994$ |

a. $\sum m = 45$

b. $\sum f = 112$

c. $\sum m^2 = 495$

d. $\sum mf = 975$

e. $\sum m^2f = 9855$

f. $\sum f^2 = 2994$