Chapter 2   Summarizing Data

I.   Linda's Video Showcase

   A.   Upon graduating from college, Linda Smith opened **Linda's Video Showcase**, a retail business specializing in videotape rentals.

   B.   Linda will use descriptive statistics to analyze this daily video rentals data set.

   1. | 76, 88, 53, 66, 97, 73, 64, 82, 77, 57, 93, 85, 70, 76, 68 |

   2.   Linda's first step was to make a list of data by order of magnitude called an **array**. She also calculated a range (high number minus the low number) for the data.

   Array: 53, 57, 64, 66, 68, 70, 73, 76, 76, 77, 82, 85, 88, 93, 97

   Range: High - Low = 97 - 53 = 44

II.   Frequency distributions

   A.   A **frequency distribution** divides data into numerical groupings and depicts the number of observations occurring within each grouping. Academic grades are often summarized with a frequency distribution with each of the five grades representing a group. A grade of B is usually between 79 and 90. The first three columns of the chart at the bottom of this page are a frequency distribution of the above rental data.

   B.   A grouping is called a **class**.
   1.   **Class limits** state the extremes of a class. Their difference is called the **class width**.
   2.   Classes must be **mutually exclusive** in that a piece of data (outcome) may belong to only one class.
   3.   Classes must be **all-inclusive (collectively exhaustive)** in that there must be a class for every outcome.

   C.   Data is often summarized with 5 to 15 classes.
   1.   A class width should be easily divisible, i.e., 5, 10, 50, 100, 500, etc.
   2.   This formula is used with a number such as five to determine an approximate class width.
   3.   Data that is naturally clustered should be so clustered in the distribution. If possible, all classes should be of equal size and contain at least one outcome.

$$\frac{range}{\# \ of \ classes} = \frac{44}{5} = 8.8$$

   D.   Rounded class limits are called **stated class limits**.
   1.   For example, a first class with stated class limits of 50-59 would have **real class limits** of 49.5-59.5.
   2.   Outcomes equal to the upper real limit belong to the next higher class. That is, the outcome 59.5 would belong to the second class.

   E.   A **tally** is a vertical line used to count class outcomes.
   1.   The total outcomes of a class are its **frequency** (rate of occurrence).
   2.   Frequency, expressed as a decimal, is called **relative frequency**.

$$relative \ frequency = \frac{class \ frequency}{total \ frequencies}$$

   F.   **Cumulative frequency** is measured by **more-than** and **less-than ogives**. Ogives summarize the cumulative number of outcomes over or under each real class limit.

   G.   Frequency, relative frequency, and cumulative frequency are calculated below and graphed on the next page.

Linda's Video Showcase Daily Rentals Beginning 1/2/98					Cumulative Frequency	
Stated Class Limits	Real Class Limits	Tally	Frequency (f)	Relative Frequency f ÷ n	More-than	Less-than
50 - 59	49.5 - 59.5	I I	2	0.13	49.5 is 15	49.5 is 0
60 - 69	59.5 - 69.5	I I I	3	0.20	59.5 is 13	59.5 is 2
70 - 79	69.5 - 79.5	⊞	5	0.34	69.5 is 10	69.5 is 5
80 - 89	79.5 - 89.5	I I I	3	0.20	79.5 is 5	79.5 is 10
90 - 99	89.5 - 99.5	I I	2	0.13	89.5 is 2	89.5 is 13
		Totals	n = 15	1.00	99.5 is 0	99.5 is 15