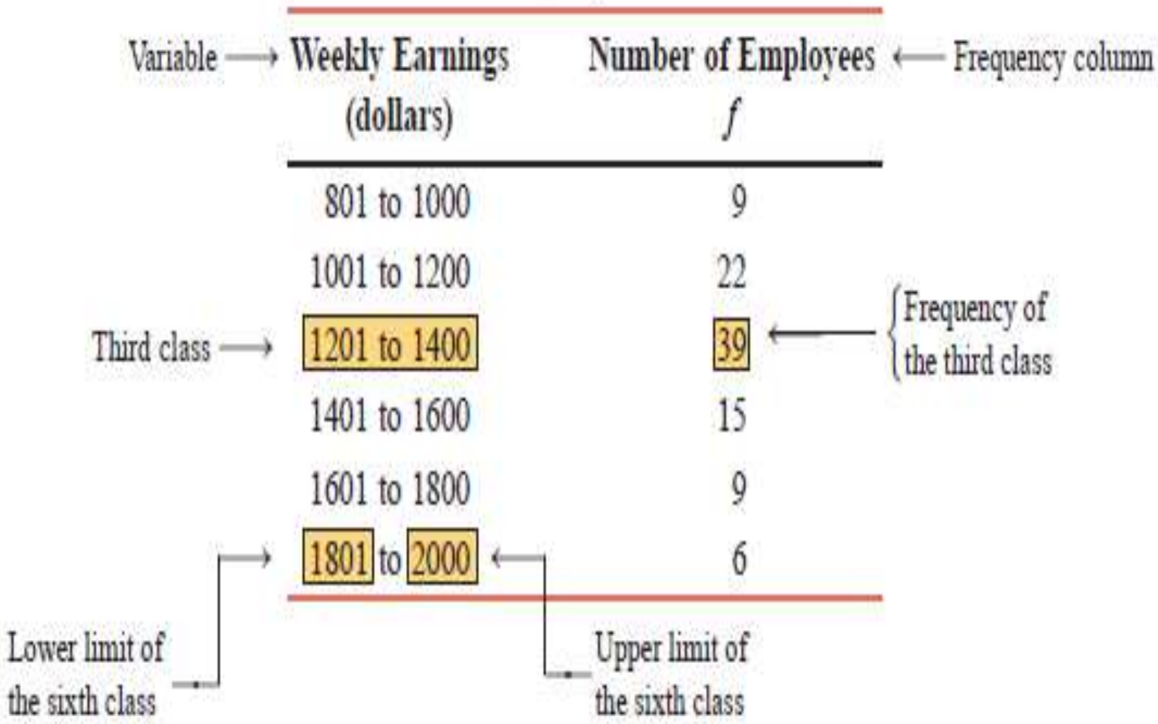
Frequency Distributions

Table (5) gives the weekly earnings of 100 employees of a large company. The first column lists the classes, which represent the (quantitative) variable weekly earnings. For quantitative data, an interval that includes all the values that fall within two numbers—the lower and upper limits—is called a class. Note that the classes always represent a variable. As we can observe, the classes are nonoverlapping; that is, each value on earnings belongs to one and only one class. The second column in the table lists the number of employees who have earnings within each class. For example, 9 employees of this company earn $801 to $1000 per week. The numbers listed in the second column are called the frequencies, which give the number of values that belong to different classes. The frequencies are denoted by *f*.



Weekly Earnings (dollars)	Number of Employees *f*
801 to 1000	9
1001 to 1200	22
1201 to 1400	39
1401 to 1600	15
1601 to 1800	9
1801 to 2000	6

For quantitative data, the frequency of a class represents the number of values in the data set that fall in that class. Table 2.7 contains six classes. Each class has a lower limit and an upper limit. The values 801, 1001, 1201, 1401, 1601, and 1801 give the lower limits, and the values 1000, 1200, 1400, 1600, 1800, and 2000 are the upper limits of the six classes, respectively.

The data presented in Table (5) are an illustration of a frequency distribution table for quantitative data. Whereas the data that list individual values are called ungrouped data, the data presented in a frequency distribution table are called grouped data.

Definition

Frequency Distribution for Quantitative Data A *frequency distribution* for quantitative data lists all the classes and the number of values that belong to each class. Data presented in the form of a frequency distribution are called *grouped data*.

To find the midpoint of the upper limit of the first class and the lower limit of the second class in Table (5), we divide the sum of these two limits by 2. Thus, this midpoint is

$$\frac{1000 + 1001}{2} = 1000.5$$

The value 1000.5 is called the *upper boundary* of the first class and the *lower boundary* of the second class. By using this technique, we can convert the class limits of Table (5) to **class boundaries**, which are also called *real class limits*. The second column of Table 2.8 lists the boundaries for Table (5).

Definition

Class Boundary The *class boundary* is given by the midpoint of the upper limit of one class and the lower limit of the next class.

The difference between the two boundaries of a class gives the **class width**. The class width is also called the **class size**.

Finding Class Width

$$\text{Class width} = \text{Upper boundary} - \text{Lower boundary}$$

Thus, in Table (6), the class widths for the frequency distribution of Table (5) are listed in the third column of Table 2.8. Each class in Table 2.8 (and Table (5) has the same width of 200.

The class midpoint or mark is obtained by dividing the sum of the two limits (or the two boundaries) of a class by 2.

Calculating Class Midpoint or Mark

$$\text{Class midpoint or mark} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Thus, the midpoint of the first class in Table (5) or Table (6) is calculated as follows:

The class midpoints for the frequency distribution of Table (5) are listed in the fourth column of Table (6).

Table (6)

Class Limits	Class Boundaries	Class Width	Class Midpoint
801 to 1000	800.5 to less than 1000.5	200	900.5
1001 to 1200	1000.5 to less than 1200.5	200	1100.5
1201 to 1400	1200.5 to less than 1400.5	200	1300.5
1401 to 1600	1400.5 to less than 1600.5	200	1500.5
1601 to 1800	1600.5 to less than 1800.5	200	1700.5
1801 to 2000	1800.5 to less than 2000.5	200	1900.5

Note that in Table 2.8, when we write classes using class boundaries, we write *to less than* to ensure that each value belongs to one and only one class. As we can see, the upper boundary of the preceding class and the lower boundary of the succeeding class are the same.

<u>Constructing Frequency Distribution Tables</u>

When constructing a frequency distribution table, we need to make the following three major decisions.

- **Number of Classes**

Usually the number of classes for a frequency distribution table varies from 5 to 20, depending mainly on the number of observations in the data set.1 It is preferable to have more classes as the size of a data set increases. The decision about the number of classes is arbitrarily made by the data organizer.

- **Class Width**

Although it is not uncommon to have classes of different sizes, most of the time it is preferable to have the same width for all classes. To determine the class width when all classes are the same size, first find the difference between the largest and the smallest values in the data. Then, the approximate width of a class is obtained by dividing this difference by the number of desired classes.

Calculation of Class Width

$$\text{Approximate class width} = \frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}$$

Usually this approximate class width is rounded to a convenient number, which is then used as the class width. Note that rounding this number may slightly change the number of classes initially intended.

- **Lower Limit of the First Class or the Starting Point**

Any convenient number that is equal to or less than the smallest value in the data set can be used as the lower limit of the first class.

Example 2–3 illustrates the procedure for constructing a frequency distribution table for quantitative data.

The following data give the total number of iPods® sold by a mail order company on each of 30 days. Construct a frequency distribution table.

8	25	11	15	29	22	10	5	17	21
22	13	26	16	18	12	9	26	20	16
23	14	19	23	20	16	27	16	21	14

Solution In these data, the minimum value is 5, and the maximum value is 29. Suppose we decide to group these data using five classes of equal width. Then,

$$\text{Approximate width of each class} = \frac{29 - 5}{5} = 4.8$$

Now we round this approximate width to a convenient number, say 5. The lower limit of the first class can be taken as 5 or any number less than 5. Suppose we take 5 as the lower limit of the first class. Then our classes will be

5–9, 10–14, 15–19, 20–24, and 25–29

We record these five classes in the first column of Table (7).

One rule to help decide on the number of classes is Sturge's formula:

$$c = 1 + 3.3 \log n$$

Where *c* is the number of classes and *n* is the number of observations in the data set. The value of log *n* can be obtained by using a calculator.

Now we read each value from the given data and mark a tally in the second column of Table (7) next to the corresponding class. The first value in our original data set is 8, which belongs to the 5–9 class. To record it, we mark a tally in the second column next to the 5–9 class.

We continue this process until all the data values have been read and entered in the tally column.

Note that tallies are marked in blocks of five for counting convenience. After the tally column is completed, we count the tally marks for each class and write those numbers in the third column.

This gives the column of frequencies. These frequencies represent the number of days on which iPods indicated in classes are sold. For example, on 8 of 30 days, 15 to 19 iPods were sold.

on iPods Sold

iPods Sold	Tally	f
5–9	|||	3
10–14	~~||||~~ |	6
15–19	~~||||~~ |||	8
20–24	~~||||~~ |||	8
25–29	~~||||~~	5
		$\Sigma f = 30$