Chapter 2: Frequency Distributions

Chapter Outline

2.1 Frequency Distributions and Frequency Distribution Tables
    Frequency Distribution Tables
    Proportions and Percentages
2.2 Grouped Frequency Distribution Tables
    Real Limits and Frequency Distributions
2.3 Frequency Distribution Graphs
    Graphs for Interval or Ratio Data
    Graphs for Nominal or Ordinal Data
    Graphs for Population Distributions
    The Shape of a Frequency Distribution
2.4 Percentiles, Percentile Ranks, and Interpolation
    Cumulative Frequency and Cumulative Percentage
    Interpolation
2.5 Stem and Leaf Displays
    Comparing Stem and Leaf Displays with Frequency Distributions

Learning Objectives and Chapter Summary

1. Describe the basic elements of a frequency distribution table and explain how they are related to the original set of scores.

The goal of descriptive statistics is to simplify the organization and presentation of data. One descriptive technique is to place the data in a frequency distribution table or graph that shows exactly how many individuals (or scores) are located in each category on the scale of measurement.

2. Calculate the following from a frequency table: ΣX, ΣX^2, and the proportion and percentage of the group associated with each score.

A frequency distribution table lists the categories that make up the scale of measurement (the *X* values) in one column. Beside each *X* value, in a second column, is the frequency (*f*) or number of individuals in that category. The table may include a proportion (*p*) column showing the relative frequency for each category and it may include a percentage column showing the percentage (%) associated with each X value.

3. Identify when it is useful to set up a grouped frequency distribution table, and explain how to construct this type of table for a set of scores.

It is recommended that a frequency distribution table have a maximum of 10–15 rows to keep it simple. If the scores cover a range that is wider than this suggested maximum, it is customary to divide the range into sections called class intervals. These intervals are then listed in the frequency distribution table along with the frequency or number of individuals with scores in each interval. The result is called a grouped frequency distribution. The guidelines for constructing a grouped frequency distribution table are as follows:

a. There should be about 10 intervals.
b. The width of each interval should be a simple number (e.g., 2, 5, or 10).
c. The bottom score in each interval should be a multiple of the width.
d. All intervals should be the same width, and they should cover the range of scores with no gaps.

4. Describe how the three types of frequency distribution graphs - histograms, polygons, and bar graphs - are constructed and identify when each is used.

5. Describe the basic elements of a frequency distribution graph and explain how they are related to the original set of scores.

6. Explain how frequency distribution graphs for populations differ from the graphs used for samples.

A frequency distribution graph lists scores on the horizontal axis and frequencies on the vertical axis. The type of graph used to display a distribution depends on the scale of measurement used. For interval or ratio scales, you should use a histogram or a polygon. For a histogram, a bar is drawn above each score so that the height of the bar corresponds to the frequency. Each bar extends to the real limits of the score, so that adjacent bars touch. For a polygon, a dot is placed above the mid- point of each score or class interval so that the height of the dot corresponds to the frequency; then lines are drawn to connect the dots. Bar graphs are used with nominal or ordinal scales. Bar graphs are similar to histograms except that gaps are left between adjacent bars.

7. Identify the shape - symmetrical, and positively or negatively skewed - of a distribution in a frequency distribution graph.

Shape is one of the basic characteristics used to describe a distribution of scores. Most distributions can be classified as either symmetrical or skewed. A skewed distribution that tails off to the right is positively skewed. If it tails off to the left, it is negatively skewed.

8. Define percentiles and percentile ranks.

9. Determine percentiles and percentile ranks for values corresponding to real limits in a frequency distribution table.

The cumulative percentage is the percentage of individuals with scores at or below a particular point in the distribution. The cumulative percentage values are associated with the upper real limits of the corresponding scores or intervals.

Percentiles and percentile ranks are used to describe the position of individual scores within a distribution. Percentile rank gives the cumulative percentage associated with a particular score. A score that is identified by its rank is called a percentile.

10. Estimate percentiles and percentile ranks using interpolation for values that do not correspond to real limits in a frequency distribution table.

When a desired percentile or percentile rank is located between two known values, it is possible to estimate the desired value using the process of interpolation. Interpolation assumes a regular linear change between the two known values.

11. Describe the basic elements of a stem and leaf display and explain how the display shows the entire distribution of scores.

A stem and leaf display is an alternative procedure for organizing data. Each score is separated into a stem (the first digit or digits) and a leaf (the last digit). The display consists of the stems listed in a column with the leaf for each score written beside its stem. A stem and leaf display is similar to a grouped frequency distribution table, however the stem and leaf display identifies the exact value of each score and the grouped frequency distribution does not.

Other Lecture Suggestions

1. Begin with an unorganized list of scores as in Example 2.1, and then organize the scores into a table. If you use a set of 20 or 25 scores, it will be easy to compute proportions and percentages for the same example.

2. Present a relatively simple, regular frequency distribution table (for example, use scores of 5, 4, 3, 2, and 1 with corresponding frequencies of 1, 3, 5, 3, 2. Ask the students to determine the values of N and ΣX for the scores. Note that ΣX can be obtained two different ways: 1) by computing and summing the fX values within the table, 2) by retrieving the complete list of individual scores and working outside the table.

Next, ask the students to determine the value of ΣX^2. You probably will find a lot of wrong answers from students who are trying to use the fX values within the table. The common mistake is to compute $(fX)^2$ and then sum these values. Note that whenever it is necessary to do complex calculations with a set of scores, the safe method is to retrieve the list of individual scores from the table before you try any computations.

3. It sometimes helps to make a distinction between graphs that are being used in a formal presentation and sketches that are used to get a quick overview of a set of data. In one case, the graphs should be drawn precisely and the axes should be labeled clearly so that the graph can be easily understood without any outside explanation. On the other hand, a sketch that is intended for your own personal use can be much less precise. As an instructor, if you are expecting

precise, detailed graphs from your students, you should be sure that they know your expectations.

4. Introduce interpolation with a simple, real-world example. For example, in Buffalo, the average snowfall during the month of February is 30 inches. Ask students, how much snow they would expect during the first half of the month. Then point out that the same interval (February) is being measured in terms of days and in terms of inches of snow. A point that is half-way through the interval in terms of days should also be half-way through the interval in terms of snow.

5. Refer to Box 2.1 The Use and Misuse of Graphs and discuss common misuses. For more examples, refer to the subtly-named [How to Lie with Data Visualization](http://data.heapanalytics.com/how-to-lie-with-data-visualization/) (http://data.heapanalytics.com/how-to-lie-with-data-visualization/). Challenge students to bring in examples of misleading graphs they find online or in print. (Hint: The more stridently a website advocates for or against a particular point of view on a social, political or other controversial issue, the more likely you are to find misrepresentation of data.)

Answers to Even-Numbered Problems

2.

X	f	p	%
9	1	0.05	5%
8	0	0.00	0%
7	1	0.05	5%
6	2	0.10	10%
5	4	0.20	20%
4	2	0.10	10%
3	3	0.15	15%
2	5	0.25	25%
1	2	0.10	10%

4. a. $n = 14$
   b. $\Sigma X = 44$
   c. $\Sigma X^2 = 168$

6.

X	*f*
60-64	1
55-59	2
50-54	2
45-49	1
40-44	2
35-39	3
30-34	3
25-29	5
20-24	8
15-19	3

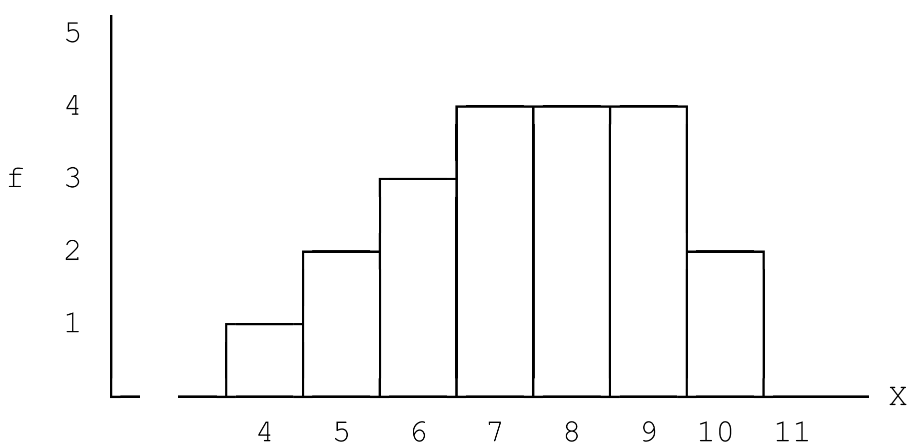
Younger drivers, especially those 20 to 29 years old, tend to get more tickets.

8. A regular table reports the exact frequency for each category on the scale of measurement. After the categories have been grouped into class intervals, the table reports only the overall frequency for the interval but does not indicate how many scores are in each of the individual categories.

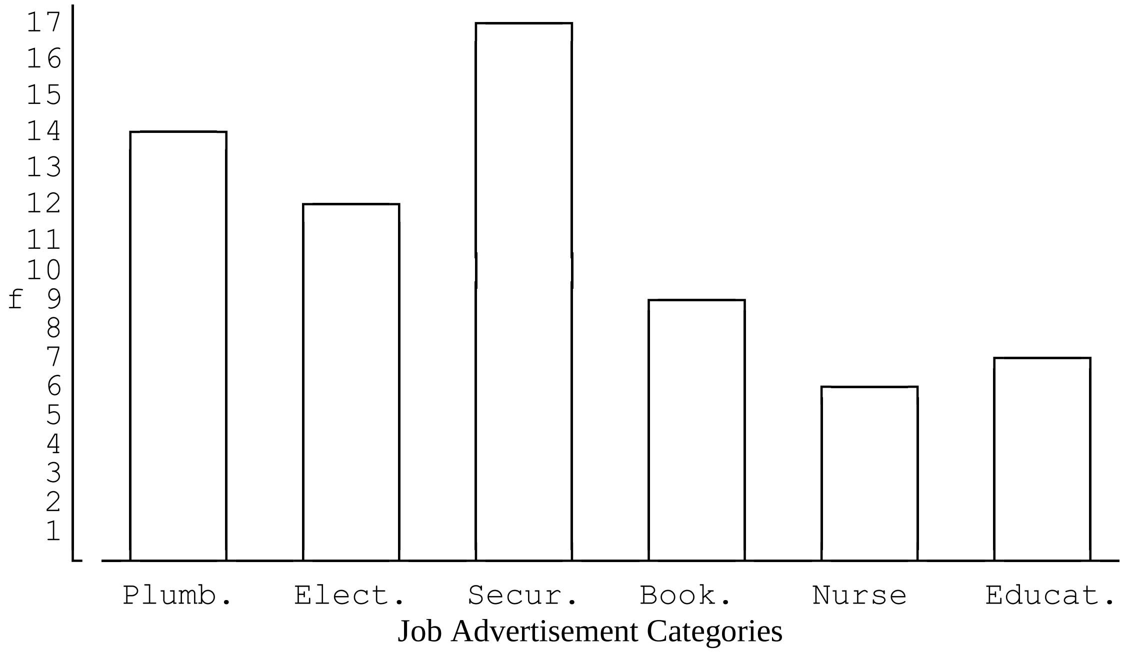
10. a.

X	*f*
10	2
9	4
8	4
7	4
6	3
5	2
4	1

b.



12.



14.

X	*f*
15	5
14	6
13	4
12	2
11	2
10	1

The distribution is negatively skewed.

16.

X	f_{low}	f_{high}
5	1	5
4	2	6
3	4	3
2	5	1
1	3	0

The scores for children from the high number-talk parents are noticeably higher.

18.

X	*f*	*cf*	*c*%
25–29	1	25	100
20–24	4	24	96
15–19	8	20	80
10–14	7	12	48
5–9	3	5	20
0–4	2	2	8

a. 20%
b. 80%
c. $X = 14.5$
d. $X = 24.5$

20. a. 32%
b. 59%
c. $X = 14.25$
d. $X = 15.10$

22. a. $X = 22$
b. $X = 31.5$
c. 93%
d. 28%

24. a. 6
b. 75, 79, 72, 73, 77, 74
c. 1
d. 48