Answers to Analyzing the Databases Questions in Chapter 2

1. Frequency Distribution of Number of Production Workers using Excel. Note in Excel if you do not specify the interval classes, it arbitrarily selects the class intervals. The last interval says ‘more’. To avoid this, the user can specify the ‘bin’. See the second table.

|  |  |
| --- | --- |
| *Bin* | *Frequency* |
| 50 | 68 |
| 100 | 30 |
| 150 | 19 |
| 200 | 10 |
| 250 | 6 |
| 300 | 1 |
| 350 | 1 |
| 400 | 2 |
| 450 | 1 |
| 500 | 0 |
| More | 2 |

0 to 50 68

51 to 100 30

101 to 150 19

151 to 200 10

201 to 250 6

251 to 300 1

301 to 350 1

351 to 400 2

401 to 450 1

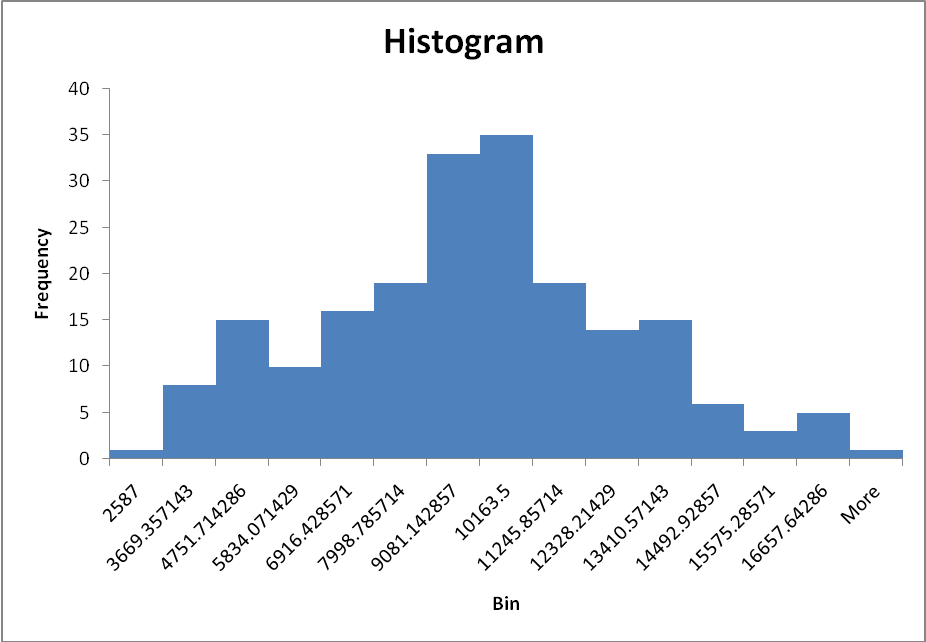
451 to 500 0

501 to 550 0

551 to 600 1

601 to 650 1

This tells us that the number of production workers “pile up” at the lower numbers. In chapter 3, we will learn that this distribution is skewed right. Number of production workers is in units of 1,000. Thus, 68 of the 140 companies in this database have between 1,000 and 50,000 workers. This represents almost 49% of the companies. An additional 30/140 or over 21% of the companies have between 50,001 and 100,000 production workers. Taken together, these two categories account for over 70% of the companies in this database.



The histogram has a somewhat symmetrical shape. It is high in the middle and “thin” at the ends. It is not unlike the normal distribution (bell-shaped curve) that will be introduced later in the text. Annual food spending seems to be centered at the $10,000 amount. A large chunk of the histogram falls between $7,000 and $13,000.

1. Graphing discrete numbers does not exactly produce a true ogive. However, one can produce a cumulative line graph over the numbers progressing from 1 to 7. The cumulative distribution is:

Type Cumulative Frequency

1. 10
2. 29
3. 42
4. 50
5. 65
6. 81

7 100

The ogive-like graph is shown below. The line in the graph is relatively straight indicating that most increases are fairly constant and there are few “jumps” or inordinate increases. The raw cumulative numbers shown above point this out. If you were to look at the actual frequency counts, you would see that Type 2 and Type 7 have the highest number with a frequency of 19 and Type 4 has the fewest with 8. Thus, as the ogive goes from Type 3 to Type 4, there is a slight dip in the increase of the line. As the ogive goes from Type 1 to Type 2 and from Type 6 to Type 7, you see the greatest slope or increase in the line.





The pie chart demonstrates that more companies are chemical and petroleum (19) than any other categories and these two categories are equal. It also shows that there are fewer grocery companies. Without the frequency labels, however, it would be more difficult to differentiate between the numbers of companies in the various categories.

4. Stem and Leaf Plot for Italy:

**Stem-and-Leaf Display: Italy**

Stem-and-leaf of Italy N = 40

Leaf Unit = 0.10

2 478

3 12234455577789

4 114489

5 499

6 09

7 035899

8

9

10 2

11 3

12 0133

The stem and leaf plot retains all of the original data so we know, for example, that some of the annual unemployment rates included 2.4%, 2.7%, 2.8%, 6.9%, 10.2%, and 12.3%. Studying this plot, we can tell that for most years (34 out of 40), the unemployment rate was below 8%. In fact, a high number of unemployment rates were in the 3% bracket. However, since these rates were gathered over a 40-year period, we cannot see from the stem and leaf plot, if there are any trends that would help predict ensuing years’ unemployment rates and/or if the high unemployment rates were many years ago or more recent.