CONTENT OVERVIEW

Rows and rows of data provide little information. For example, below are thickness measurements, in millimeters, from a sample of 25 polished wafers used in the manufacture of microchips. Notice that it is difficult to extract much information from staring at these numbers. The numbers need to be organized, summarized, and displayed graphically in order to unlock the information they contain.

0.402	0.496	0.533	0.387	0.384
0.528	0.411	0.367	0.462	0.499
0.539	0.546	0.425	0.457	0.586
0.558	0.588	0.425	0.437	0.479
0.427	0.485	0.443	0.441	0.658

A **frequency distribution** is one method of organizing and summarizing data in a table. The basic idea behind a frequency distribution is to set up categories (class intervals), classify data values into the categories, and then determine the frequency with which data values are placed into each category. The steps below outline the process of making a frequency distribution table.

Creating a frequency distribution table

Step 1: Identify an interval that is wide enough to contain all the data.

Step 2: Subdivide the interval identified in Step 1 into class intervals of equal width. The class intervals will serve as the categories.

Step 3: Set up a table with three columns for the following: class interval, tally, and frequency. (The tally column can be removed in the final table.)

Step 4: To complete the table, determine the frequency with which data values fall into each class interval.

Convention: Any data value that falls on a class interval boundary is placed in the class interval to the right. If the data value is a maximum, it is generally put in the interval that contains the maximum at its right endpoint.

Now, we apply Steps 1 – 4 to make a frequency distribution table for the thickness measurements.

Step 1: In this case the smallest data value is 0.367 mm and the largest is 0.698 mm. We choose the interval from 0.3 mm to 0.7 mm, which contains all the thickness measurements.

Step 2: The total width of the interval from 0.3 to 0.7 is 0.4. Dividing this interval into eight class intervals works out nicely – each class interval will have width 0.05.

Step 3: We have set up Table 3.1 to have three columns, which we have labeled Thickness, Tally, and Frequency. We have entered the endpoints of the eight class intervals into the Thickness column.

Thickness (mm)	Tally	Frequency
0.30 – 0.35		
0.35 – 0.40		
0.40 – 0.45		
0.45 – 0.50		
0.50 – 0.55		
0.55 – 0.60		
0.60 – 0.65		
0.65 – 0.70		

Table 3.1: Setting up a frequency distribution table.

Step 4: The easiest way to determine the frequencies is to draw a tally line for each data value that falls into a particular class interval. When drawing tally lines, keep the following in mind:

- As you draw tally lines, instead of drawing a fifth tally line, cross out the previous four.
- If a data value falls on the boundary of a class interval, record it in the interval with the larger values.

Once a tally line has been drawn for each data value, count the number of tally lines corresponding to each class interval and record that number in the frequency column as shown in Table 3.2.

Thickness (mm)	Tally	Frequency
0.30 – 0.35		0
0.35 – 0.40	|||	3
0.40 – 0.45	卌 |||	8
0.45 – 0.50	卌		6
0.50 – 0.55	|||||	4
0.55 – 0.60	|||	3
0.60 – 0.65		0
0.65 – 0.70			1

Table 3.2: A completed frequency distribution table.

The frequency distribution in Table 3.2 reveals more information about the data than a quick look at the 25 numbers. For example, from the frequency distribution, we learn that more measurements fall in the interval 0.40 – 0.45 than in any of the other class intervals. Also, we learn there is a gap in the data – no data values fall between 0.60 and 0.65.

Although a frequency distribution table is a useful tool for extracting information from data, a histogram can often convey the same information more effectively. Next, we outline how to construct a histogram from a frequency distribution.

Creating a histogram from a frequency distribution

Step 1: Draw a set of axes. On the horizontal axis, mark the boundaries of the class intervals. On the vertical axis, set up a scale appropriate for the frequencies. (Later this scale can be changed to proportion or percent.)

Step 2: Label the horizontal axis with the name of the variable being measured and the units.

Step 3: Over each interval, draw a rectangle with the interval as its base. The height of the rectangle should match the frequency of data contained in that interval.

Next, we apply Steps 1 – 3 for creating a histogram to the frequency distribution in Table 3.2. Figure 3.9 shows the results.
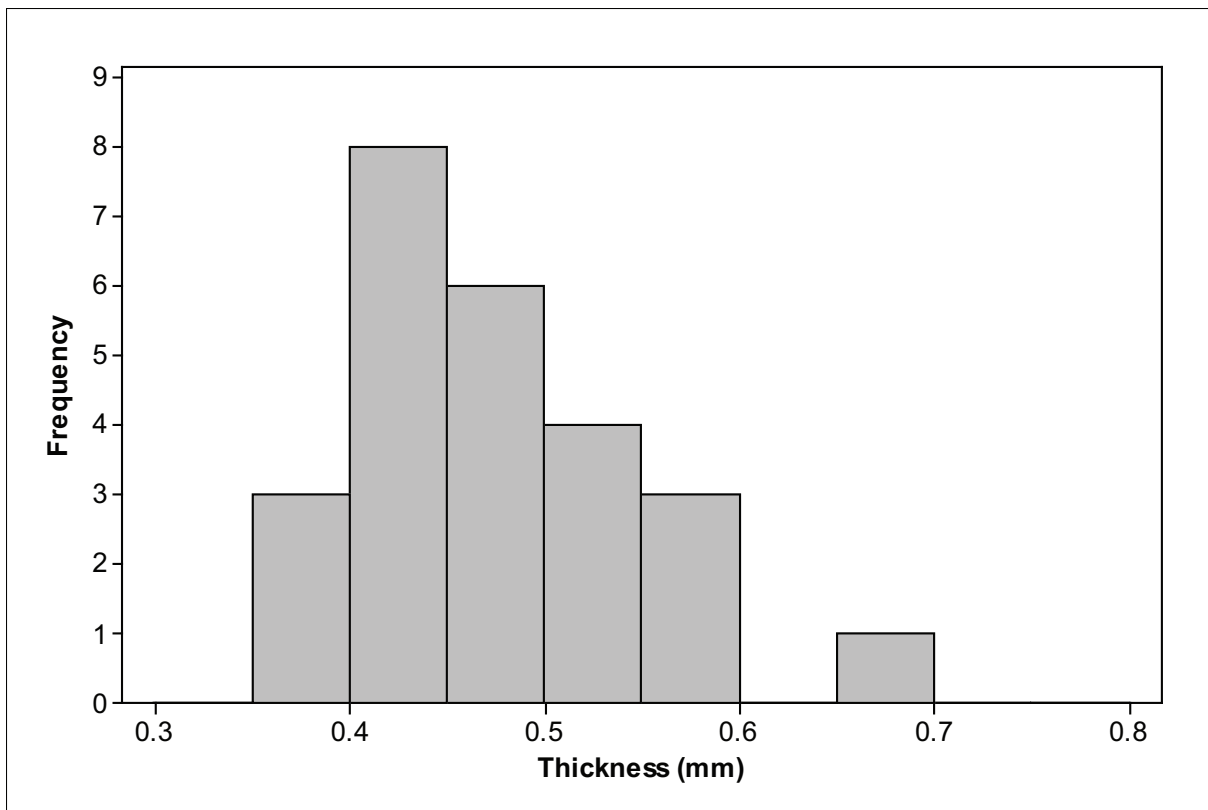
Figure 3.9: Histogram representing frequency distribution in Table 3.2.

Particularly if you are comparing histograms from samples with a different number of data values, it is useful to replace the frequency scale on the vertical axis with the proportion or percent.

Calculating Proportions and Percents

- To calculate a proportion, divide the frequency by the sample size.
- To convert a proportion into a percent, multiply the proportion by 100%.

In describing a histogram, we first look for the overall pattern of the distribution. In sizing up the overall pattern, look for the following:

- center and spread;
- one peak or several (unimodal or multimodal);
- a regular shape, such as symmetric or skewed.

In the case of the histogram in Figure 3.9, the overall pattern is single-peaked (or unimodal) and skewed to the right. Next, we look for any striking deviations from that pattern. An important kind of deviation from an overall pattern is an outlier, an individual observation

that lies clearly outside the overall pattern. Once identified, outliers should be investigated. Sometimes they are errors in the data and sometimes they have interesting stories related to the data. For Figure 3.9, there is a gap between 0.6 and 0.65 and there is one data value between 0.65 and 0.70, which might be an outlier.