

Chapter 2 Notes:

When working with large data sets, a **frequency distribution** (or **frequency table**) is often helpful in organizing and summarizing data. A frequency distribution helps us to understand the nature of the **distribution** of a data set.

Frequency Distribution (or Frequency Table): Shows how data are partitioned among several categories (or **classes**) by listing the categories along with the number (frequency) of data values in each of them.

Class midpoints: The values in the middle of the classes Each class midpoint can be found by adding the lower class limit to the upper class limit and dividing the sum by 2.

Class width: The difference between two consecutive lower-class limits in a frequency distribution.

Example. Using the McDonald's lunch service times, construct the frequency distribution:

Time(secs) 107 139 197 209 281 254 163 150 127 308 206 187 169 83 127 133 140 143 130 144 91 113 153
255 252 200 117 167 148 184 123 153 155 154 100 117 101 138 186 196 90 144 119 135 151 197 171 190 169

Time (Seconds) frequency table:

Time (secs)	Frequency
75-124	11
125-174	24
175-224	10
225-274	3
275-324	2

Class width: $125 - 75 = 50$

First class limits: lower limit, 75; upper limit, 124. First class boundaries: 74.5, 124.5

Relative frequency distribution:

Time(secs)	Relative frequency
75-124	0.22
125-174	0.48
175-224	0.20
225-274	0.06
275-324	0.04

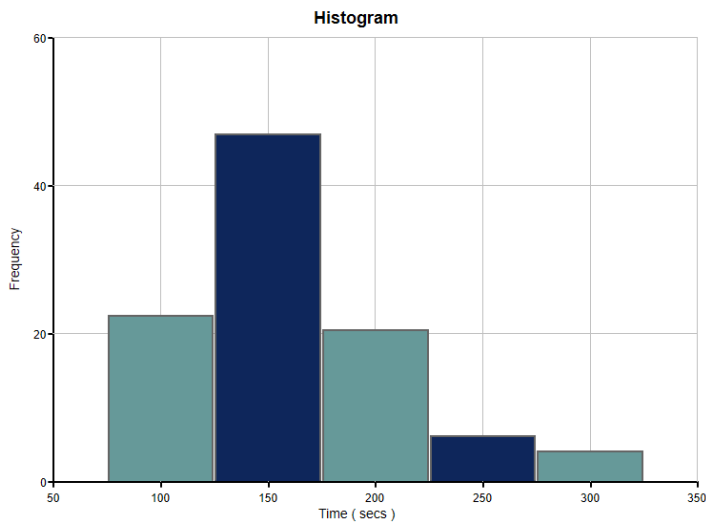
Cumulative frequency distribution:

Time(secs)	Cumulative frequency
Less than 125 s	11
Less than 175 s	35
Less than 225 s	45
Less than 275 s	48
Less than 325 s	50

Histogram: A graph consisting of bars of equal width drawn adjacent to each other (unless there are gaps in the data)

The horizontal scale represents classes of quantitative data values, and the vertical scale represents frequencies. The heights of the bars correspond to frequency values.

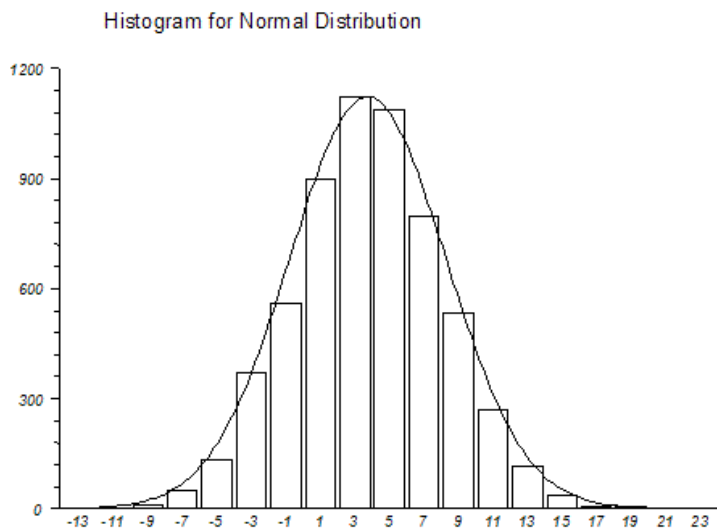
Histogram generated by statdisk (McDonald's lunch service times):



In statistics we are often interested in determining whether the data have a **normal distribution**.

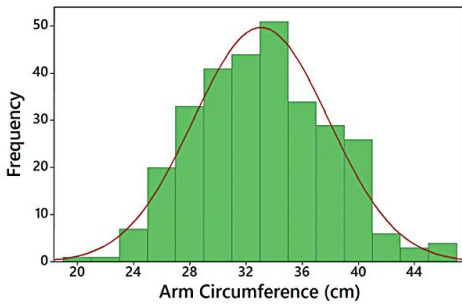
The frequencies start low, then increase to one or two high frequencies, and then decrease to a low frequency. The distribution is approximately symmetric. Frequencies preceding the maximum frequency should be roughly a mirror image of those that follow the maximum frequency.

A normal histogram looks approximately like this:



You may notice that the one we obtained for the McDonald's lunch service times example, does not follow a normal distribution.

This other example, the histogram is roughly bell-shaped; so we say that the data have a **normal distribution**.



Stemplots (or stem-and-leaf plot)

A stem & leaf plot organizes data points by the place value of the leading digits. When making a stem & leaf plot, each item of data is separated into two parts. The “stems” usually consist of the digits in the greatest common place value of each item of data. The “leaves” contain the other digits of each item of data.

Example:

For the data set: 72, 49, 62, 58, 73, 55, 78, 83, 57, 63, 73, 73, 75, 85, 85, 64, 61, 67, 75, 91.

Stem	Leaf
4	9
5	5 7 8
6	1 2 3 4 7
7	2 3 3 3 5 5 8
8	3 5 5
9	1