

Section 2.1 - Frequency Distributions

A **frequency distribution** is a table that shows how a data set is partitioned among all of several classes by listing all of the classes along with the number of data values in each class.

Example

{5.1,4.8,5.2,6.2,4.9,5.3,5.6,5.7,6.0,4.2,5.7,5.7,5.8,5.0,5.1}

Height (ft)	Frequency	Cumulative freq.
4.0-4.4	1	1
4.5-4.9	2	3
5.0-5.4	5	8
5.5-5.9	5	13
6.0-6.4	2	15

Lower class limits - smallest numbers that belong to the different classes

Upper class limits - largest numbers that belong to the different classes

Class boundaries - numbers used to separate classes; these numbers are the midpoints of the intervals from consecutive upper to lower class limits; class boundaries should not be data values

Class midpoints - values in the middle of the classes

Class width - difference between consecutive lower class limits or upper class limits

Constructing a Frequency Distribution...

Choose the number of classes, class limits, and class width in some convenient or intuitive way, or follow this procedure

1. Determine the number of classes---a convenient number between 5 and 20
2. Class width should be about $(\text{max} - \text{min})/(\text{no. of classes})$
3. Choose min data value, or convenient smaller value, for first lower class limit
4. Use the class width to find all lower class limits
5. Determine the upper class limits

Now create your list, but first use tally marks!

Example with some sunspot numbers

162.2	166.2	169.4	170.3	168.3	168.1	169.2	168.0	165.6	163.7	161.9	161.2
157.8	153.0	153.7	156.6	157.7	158.9	160.8	162.8	163.3	163.2	164.8	163.9
162.7	163.9	162.5	159.6	157.9	155.2	151.6	147.6	143.4	139.3	134.1	131.0
129.9	127.6	123.4	119.7	117.4	114.9	112.0	110.4	109.9	108.7	107.4	105.6

- **Relative frequency distributions** use percentages instead of frequency counts. The sum of the relative frequencies should be 100% (or very close to it).
- **Cumulative frequency distributions** accumulate the sums of the frequency counts at each class. The final sum must equal the number of data values.

A particular type of distribution that arises frequently in applications is the **normal distribution**.

We will study normal distributions in great detail, but for now...

A normal distribution is characterized by

- The frequencies start low, then increase to one or two high frequencies, then decrease to low frequencies
- The distribution is approximately symmetric