Section 2.1, Frequency Distributions and Their Graphs

The main characteristics we will use to describe a data set are its center, its variability, and its shape. One way to see patterns in data is to make a graph. In this section, we will look at 3 ways to graphically summarize data: frequency distributions, frequency histograms, and a cumulative frequency graph.

1 Frequency distribution

A **frequency distribution** is a table that shows "classes" or "intervals" of data entries with a count of the number of entries in each class. The **frequency** f of a class is the number of data entries in the class. Each class will have a "lower class limit" and an "upper class limit" which are the lowest and highest numbers in each class. The "class width" is the distance between the lower limits of consecutive classes. The **range** is the difference between the maximum and minimum data entries.

Steps for constructing a frequency distribution from a data set

1. If the number of classes is not given, decide on a number of classes to use. This number should be between 5 and 20.

2. Find the class width: Determine the range of the data and divide this by the number of classes. Round up to the next convenient number (if it's a whole number, also round up to the next whole number).

3. Find the class limits: You can use the minimum data entry as the lower limit of the first class. To get the lower limit of the next class, add the class width. Continue until you reach the last class. Then find the upper limits of each class (since the classes cannot overlap, and occasionally your data will include decimal numbers, remember that it's fine for the upper limits to be decimals).

4. Count the number of data entries for each class, and record the number in the row of the table for that class. (The book recommends using "tally" marks to count)

Example

Make a frequency distribution for the following data, using 5 classes:

| 5 | 10 | 7 | 19 | 25 | 12 | 15 | 7 | 6 | 8 |
| 17 | 17 | 22 | 21 | 7 | 7 | 24 | 5 | 6 | 5 |

The smallest number is 5, and the largest is 25, so the range is 20. The class width will be $20/5 = 4$, but we need to round up, so we will use 5. Our classes will be 5–9, 10–14, 15–19, 20–24, and 25–29. Then, counting the number of entries in each class, we get:

Class	Frequency
5–9	10
10–14	2
15–19	4
20–24	3
25–29	1

Note that the sum of the frequencies is 20, which is the same as number of data entries that we had.

You can add more information to your frequency distribution table. The "midpoint" (or "class mark") of each class can be calculated as:

$$\text{Midpoint} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}.$$

The "relative frequency" of each class is the proportion of the data that falls in that class. It can be calculated for a data set of size n by:

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}.$$

The "cumulative frequency" is the sum of the frequencies of that class and all previous classes.

Example

Add the midpoint of each class, the relative frequency, and the cumulative frequency to previous frequency table.

Class	Frequency	Midpoint	Relative frequency	Cumulative frequency
5–9	10	7	0.5	10
10–14	2	12	0.1	12
15–19	4	17	0.2	16
20–24	3	22	0.15	19
25–29	1	27	0.05	20

2 Frequency histogram

A **frequency histogram** is a graphical way to summarize a frequency distribution. It is a bar graph with the following properties:

1. The horizontal scale is quantitative and measures the data values.

2. The vertical scale measures the frequencies of the classes.

3. Consecutive bars must touch. As a result, the "class boundaries" are the numbers that separate classes without forming gaps. They will be the lower limits of classes as calculated for a frequency distribution.

Example

Construct a frequency histogram for the data considered before.

Done in class

3 Cumulative frequency graph

A **cumulative frequency graph**, or **ogive** is a line graph displaying the cumulative frequency of each class at its upper class boundary. The upper boundaries are marked on the horizontal axis, and the cumulative frequencies are marked on the vertical axis. The graph should start at (or just before) the lower boundary of the first class (where the cumulative frequency is zero), and end at the upper boundary of the last class. The graph should be increasing from left to right, and the points should be evenly spaced along the horizontal axis.

Example

Construct an ogive for the data considered before.

Done in class