

1.3.5.3: Grouping Numeric Data

Another way to organize raw data is to group them into **class intervals**, and to then create a **frequency distribution** of these class intervals.

There are many methods of creating class intervals, so we will simply focus on creating intervals of equal width.

How to create class intervals of equal width and a frequency distribution

1. Choose how many intervals you want. Best is between 5 and 15 intervals.
2. Determine the interval width using the formula and rounding UP to a convenient value:

$$IW = \text{Interval Width} = \frac{\text{Maximum Value} - \text{Minimum Value} + 1}{\text{Number of Intervals}}$$

3. Create the **class intervals** starting with the minimum value:

Min to under Min + IW,

Min +IW to under Min +2(IW), ...

4. Calculate the **frequency** of each class interval by counting the values in each class interval. Values that are on an endpoint should be put in the lower class interval. This result is called a **frequency distribution**.

Example: Students browsing the web

Let's return to the data that represents how much time 30 students spent on a web browser in a 24 hour period. Data is rounded to the nearest minute.

102	104	85	67	101
71	116	107	99	82
103	97	105	103	95
105	99	86	87	100
109	108	118	87	125
124	112	122	78	92

First we choose how many class intervals. In this example, we will create 5 class intervals.

Next Determine the Class Interval Width and round up to a convenient value.

$$IW = \frac{125 - 67 + 1}{5} = 11.8 \rightarrow 12$$

Now create class intervals of width 12, starting with the lowest value, 67.

(67 to 79) (79 to 91) (91 to 103) (103 to 115) (115 to 127)

Now, create a frequency distribution, by counting how many are in each interval. Values that are on an endpoint should be put in the higher class interval. For example, 103 should be counted in the interval (103 to 115):

Class Interval	Frequency
67 to 79	3
79 to 91	5
91 to 103	8
103 to 115	9
115 to 127	5
Total	30

As we did with categorical data, we can define **Relative Frequency** as the proportion or percentage of values in any Class Interval.

n = sample size - The number of observations in your sample size.

Frequency - the number of times a particular value is observed in a class interval.

Relative frequency - The proportion or percentage of times a particular value is observed in a class interval.

Relative Frequency = Frequency / n

Class Interval	Frequency	Relative Frequency
67 to 79	3	0.100 or 10.0%
79 to 91	5	0.167 or 16.7%
91 to 103	8	0.266 or 26.6%
103 to 115	9	0.300 or 30.0%
115 to 127	5	0.167 or 16.7%
Total	30	1.000 or 100%

Note that the value for the (91 to 103) class interval was deliberately rounded down to make the totals add up to exactly 100%

From the frequency distribution, we can see that 30% of the students are on the internet between 103 and 115 minutes per day, while only 10% of students are on the internet between 67 and 79 minutes.

Example: Comparing weights of apples and oranges

A Chilean agricultural researcher collected a sample of 100 Royal Gala apples and 100 navel oranges and measured their weights in grams (see previous example on dot plots).

We will start with a value of 100 and make the interval width equal to 30. Using the tally feature of Minitab, we can create a frequency distribution for the two fruits. Minitab uses “Count” for “Frequency” and reports “Percent” for “Relative Frequency”

Class interval	Apples Count	Apples Percent	Oranges Count	Oranges Percent
100 to 130	1	1.00	1	1.00
130 to 160	3	3.00	0	0.00
160 to 190	9	9.00	6	6.00
190 to 220	15	15.00	10	10.00
220 to 250	23	23.00	14	14.00
250 to 280	18	18.00	18	18.00
280 to 310	16	16.00	19	19.00
310 to 340	11	11.00	9	9.00
340 to 370	3	3.00	13	13.00
370 to 400	0	0.00	4	4.00
400 to 430	0	0.00	4	4.00
430 to 460	1	1.00	2	2.00
Totals	100	100.00	100	100.00

The most frequently occurring interval for apples is 220 to 250 grams while the most frequently occurring interval for oranges is 280 to 310 grams. Notice that there are some intervals with 0 observations, showing a potential high outlier for apples and a low outlier for oranges.

1.3.5.3: Grouping Numeric Data is shared under a [CC BY-SA](#) license and was authored, remixed, and/or curated by LibreTexts.

- 2.5.3: Grouping Numeric Data by Maurice A. Geraghty is licensed [CC BY-SA 4.0](#). Original source: <http://nebula2.deanza.edu/~mo/holisticInference.html>.