CSC1311: STATISTICS FOR PHYSICAL SCIENCE AND ENGINEERING

Lecture 2: Frequency Distribution and Graphical Representation

December 1, 2021

Frequency Distribution

- A Frequency Distribution is a grouping of data into mutually exclusive and exhaustive classes showing the number of observations in each class.

Example 1

Consider the guessed weights (lbm) collected in our first class on Nov. 24, 2021 from 62 students (Hypothetical).

140 135 140 160 175 150 152 155 155 165 145 150 154 160 143
160 170 155 140 160 160 175 140 145 150 150 152 159 160 165
145 155 150 150 165 148 152 155 155 160 172 180 141 147 155
165 170 160 140 150 150 152 155 130 155 163 170 139 165 180
180 190

Problem: Let us organize it into a frequency distribution table.

Five steps procedure to construct a frequency distribution

- **Step 1.** Decide how many classes you wish to use. Step 2. Determine the class width
- **Step 3.** Set up the individual class limits
- **Step 4.** Tally the items into the classes
- **Step 5.** Count the number of items in each class

Step 1. Decide how many classes you wish to use

- Rule of Thumb: Use the 2 to the kth rule.
- Suppose there are n points in the data: Choose k so that 2 raised to the power of k is greater than n; namely $k \geq log_2^n$.
- For this example, $n = 62$, so $k = 6$ because $26 = 64 \geq 62$; or $log_2^{62} \approx 5.954196 \approx 6$.

Step 2. Determine the class width

- Generally, the class width should be the same size for all classes.
- $$C = \left\lceil \frac{max - min}{k} \right\rceil$$
- For this example,

$$C = \left\lceil \frac{190 - 130}{6} \right\rceil = 10$$

Step 3. Set up the individual class limits

It is the science and art

- We only need to know the lower limit of the first class L.
- $$L = \left\lceil min - \frac{C * k - (max - min)}{2} \right\rceil$$
- For this example,

$$L = \left\lceil 130 - \frac{10 * 6 - (190 - 130)}{2} \right\rceil = 130$$

Frequency Distribution Table for the weight example

- Solution: Tallying and counting in Steps 4 and 5 result in the following frequency distribution table.

class	frequency
$[130, 140)$	3
$[140, 150)$	12
$[150, 160)$	23
$[160, 170)$	14
$[170, 180)$	6
$[180, 190]$	4

Table: Frequency Table

Some terminologies associated with the table

- Data organized into a frequency distribution table also called grouped data.
- Class frequency: The number of observations in each class.
- Class relative frequency: The percent of observations in each class.
- Class cumulative frequency: The total observations up to certain class
- Class Midpoint: A point that divides a class into two equal parts, i.e. the average of the upper and lower class limits.
- Class interval (a.k.a. class width or class size): The class interval is obtained by subtracting the lower limit of a class from the lower limit of the next class.

Terminologies associated with the table

class	frequency	relative freq.	cummulative freq.	mid point
$[130, 140)$	3	0.05	3	135
$[140, 150)$	12		15	
$[150, 160)$	23			
$[160, 170)$	14			
$[170, 180)$	6			
$[180, 19]$	4			

Table: Frequency Table

Histogram

- A Histogram is a graph in which the classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars and the bars are drawn adjacent to each other.

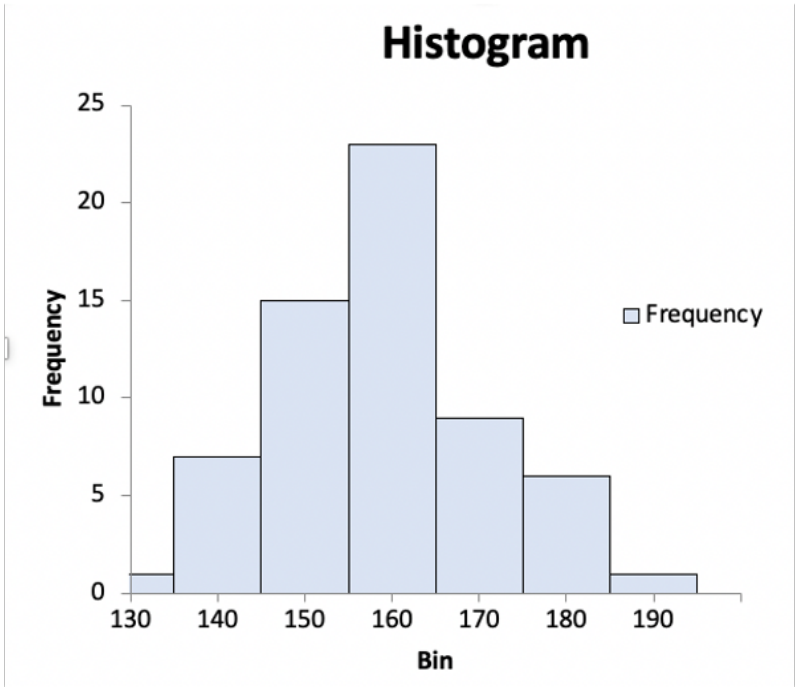

Figure: Histogram

Polygon

- A frequency polygon consists of line segments connecting the points formed by the class midpoint and the class frequency.

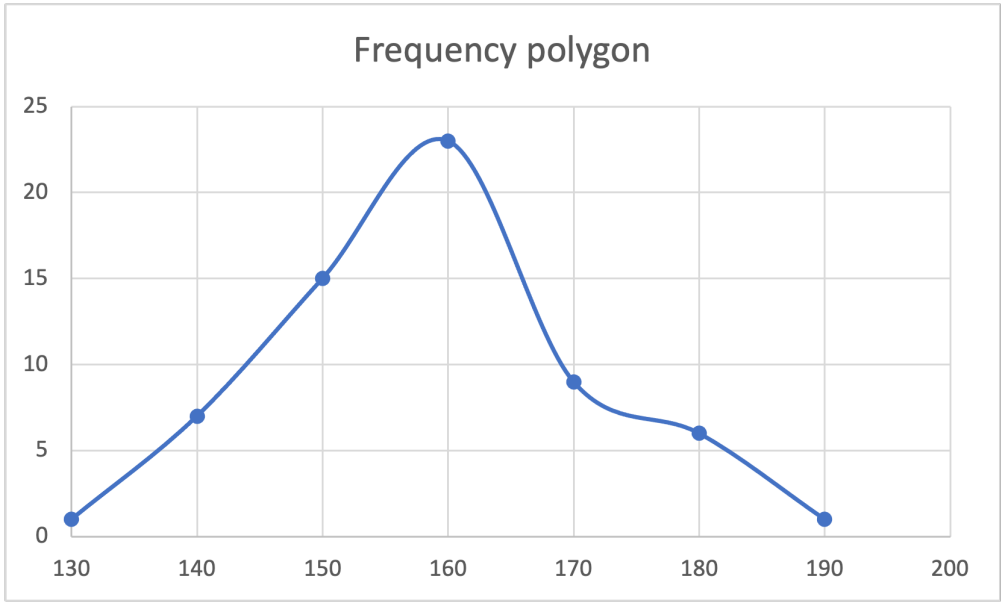

Figure: Histogram

Ogive: cumulative frequency polygon

- An ogive consists of line segments connecting the points formed by the class upper limits and the class frequency.
- A cumulative frequency polygon is used to determine how many or what proportion of the data values are below or above a certain value.

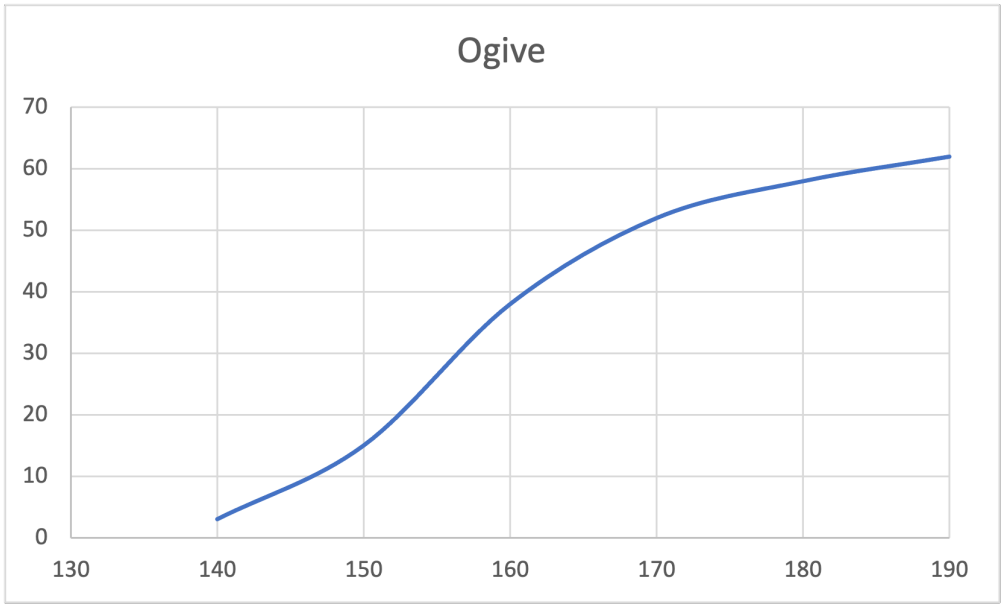

Figure: Histogram

Stem-and-leaf display

- A statistical technique for displaying a set of data, and each numerical value is divided into two parts:
  - the leading digits become the stem
  - the trailing digits become the leaf.
- One advantage of the stem-and-leaf display over a frequency distribution is that we retain the value of each observation!
- Another is the distribution of the data within each groups is clear.

How to develop a stem-and-leaf display

- **Step 1:** (Identify the stem) This can be done as follows:
  - Find the lowest value, record the leading digit.
  - Find the next score with the second highest leading digit.
  - Repeat the above until all data are examined
- Step 2: (Identify the leaf) list the remaining leaf values based on the stems.

Draw the stem-and-leaf display for the example

- The stem-and-leaf display for the weight example
- The decimal point is 1 digit(s) to the right of the —: