# **Chapter 2. Summarizing and Graphing Data**

**2-2**

When data are collected in original form, they are called ***raw data***.

The ***frequency*** is the number of values in a specific class of the distribution.

A ***frequency distribution*** is the organization of raw data in table form, using classes and frequencies.

Three types of frequency distributions:

1. **Categorical frequency distribution**

The categorical frequency distribution is used for data that can be placed in specific categories, such as nominal-level data.

For example, data such as political affiliation, religious affiliation, or major field of study would use categorical frequency distribution.

**Constructing a categorical frequency distribution**:

Step1.Make a table with following headings

Class Tally Frequency cf

Step2. Tally the data and place the results in tally column.

Step 3. Count the tallies and place the results in frequency column.

 Step 4. Find cumulative frequency (c.f.) for each class.

 Step 5. Find the totals for columns of frequency and c.f..

Twenty-five army inductees were given a blood test to determine their blood type. The data set is

A B B AB O

O O B AB B

B B O A O

A O O O AB

AB A O B A

Construct a frequency distribution for the data.

**2. Grouped Frequency Distribution**

When the *range of the data* is large the data must be grouped into classes that are more than one unit in width.

For example, a distribution of the number of hours that boat batteries lasted is:

**Class class Tally Frequency CF**

**Limits boundaries**

24-30 23.5-30.5 /// 3 3

31-37 30.5-37.5 / 1 4

38-44 37.5-44.5 ///// 5 9

45-51 44.5-51.5 ///// //// 9 18

52-58 51.5-58.5 ///// / 6 24

59-65 58.5-65.5 / 1 25

***Class limits*:**

The values 24 and 30 of the first class are called class limits. The **lower class limit** is 24—it represents the smallest data value that can be included in the class.

The **upper class limit** is 30—it represents the largest data value that can be included in the class.

***Class boundaries:***

The numbers in the second column are called **class boundaries.** These numbers are used to separate the classes so that there are no gaps in the frequency distribution (there is a gap between 30 and 31).

* Class limits should have the same decimal place value as the data but the class boundaries should have one additional place value and end in a 5.
* Find the boundaries by subtracting 0.5 from lower class limit and adding 0.5 to upper class limit.

**Lower limit - 0.5 = lower boundary**

**Upper limit + 0.5 = upper boundary**

* If the data are in tenths, such as 6.2, 7.8 and 12.6, the limits for a class might be 7.8-8.8, and the boundaries for that class would be 7.75-8.85, found by subtracting 0.05 from 7.8 and adding 0.05 to 8.8.

***Class width:***

The class width for a class is found by subtracting the lower boundary from the upper boundary for any given class.

 

\*Don’t subtract the limits of a single class. It will result in an incorrect answer.

Follow these rules to construct a frequency distribution:

1. There should be between 5 and 20 classes.
2. The class width should be an odd number. This ensures that the midpoint (Xm) of each class has the same place value as the data.

 

or



Midpoints are necessary for graphing.

1. The classes must be mutually exclusive i.e. they should have nonoverlapping class limits so that data cannot be placed into two classes.
2. The classes must be continuous. There should be no gaps in a frequency distribution.
3. The classes must be exhaustive. There should be enough classes to accommodate all the data.
4. The classes must be equal in width (except when there is an open-ended distribution).

**Constructing a Grouped Frequency Distribution:**

Step1. Determine the classes

 a. Find the highest and lowest value.

 b. Find the range. Range = highest value – lowest value

 c. Select the number of classes desired.

 d. Find the class width.

 e. Select a starting point (usually the lowest value) and

 add the width to get next lower limits.

 f. Find the upper class limit by subtracting 1 from the

 lower limit of the second class. Then add class width

 to each upper limit to get all the upper limits.

 g. Find the boundaries (by subtracting 0.5 from each

 lower class limit and adding 0.5 to each upper class

 limit.

Step2. Tally the data

Step 3. Find the numerical frequencies from the tallies.

Step4. Find the cumulative frequencies.

**3.Ungrouped Frequency Distribution:**

When the range of the data values is relatively small, a frequency distribution can be constructed using single data values for each class.

**2-3 Histograms**

**Histograms:**

The graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

**Constructing a histogram**:

Step1. Write down class boundaries (if not already given).

Step2. Draw and label the x and y axes.

Step3. Represent the frequency on the y-axis and the ***class boundaries*** on the x-axis.

Step 4. Using the frequencies as the heights, draw vertical bars for each class.

**The Frequency Polygon:**

The graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes.

**Constructing a frequency polygon:**

Step1. Find the midpoints of each class.

 

or

 

 Step 2. Draw the x and y axes. Label the x-axis with the midpoint of each class, and then use a suitable scale on the y-axis for the frequency.

Step 3. Using the midpoints for the x values and the frequencies as the y values, plot the points.

Step4. Connect adjacent points with line segments. Draw a line back to the x-axis at the beginning and end of the graph, at the same distance that the previous and next midpoints would be located.

**The Cumulative Frequency graph or Ogive:**

The graph that represents the cumulativefrequencies for the classes in a frequency distribution.

**Constructing an ogive:**

Step1. Find the cumulative frequency for each class.

Step2. Draw the x and y-axes. Label the x axis with the

class boundaries. Use an appropriate scale for the y-axis to represent the cumulative frequencies.

Step3. Plot the cumulative frequency at each upper class boundary. (Upper boundaries are used since the cumulative frequencies represent the number of data values accumulated up to the upper boundary.

Step4. Starting with the first upper class boundary, connect adjacent points with line segments. Then extend the graph to the first lower class boundary on the x-axis.

\* Cumulative frequency graphs are used to visually represent how many values are below a certain upper class boundary.

**Distribution Shapes:**

A frequency distribution can have many shapes. Some of the most common are:

1. **Bell shaped distribution** has a single peak and tapers off at either end. It is approximately symmetric.
2. **Uniform distribution** is basically flat or rectangular.
3. **J-shaped distribution** has a few data values on the left side and increases as one moves to the right. A reverse J-shaped distribution is the opposite of the J-shaped distribution.
4. **Right skewed distribution** has the peak of the distribution to the left and the data values taper off to the right.
5. **Left skewed distribution** has the peak of the distribution to the right and the data values taper off to the left.
6. **Unimodal distribution** has one peak
7. **Bimodal distribution** has two peaks of the same height.
8. **U-shaped distribution**

**2-4 Other types of Graphs**

1. **Pareto Chart:** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

**Constructing a Pareto chart:**

Step1. Arrange the data from the largest to smallest according to frequency.

Step2. Draw and label the x and y axes.

Step3. Draw the bars corresponding to the frequencies.

**2.The time Series Graph**: represents data that occur over a specific period of time.

**Constructing a time series graph**:

Step1. Draw and label the x and y axes.

Step2. Label the x axis for years and the y axis for the number of item.

Step3. Plot each point according to the table.

Step4. Draw line segments connecting adjacent points.

1. **The Pie Graph:** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

**Constructing a pie graph:**

Step1. Convert the frequency of each class into a proportional part of the circle by using the formula

 

where f = frequency for each class and n = sum of the frequencies.

\* The degrees should sum to 360o

\* The degrees column does not always sum to 3600 due to rounding and the percent column does not always sum to 100% for the same reason.

Step2. Convert each frequency to a percentage using the formula

 

Step3. Using a protractor, a compass and the appropriate degree measures found in step1, draw the graph and label each section with the name and percentage.

1. **Stem and Leaf Plots:**

A **stem and leaf plot** is a data plot that uses part of the data value as the stem and part of the data value as the leaf to form groups or classes.

**Constructing a stem and leaf plot:**

* 1. Arrange the data in order.
	2. Separate the data according to the first digit.
	3. Make a table using the leading digit as the stem and the trailing digit as the leaf.

**Leading digit (stem) Trailing digit (leaf)**

* 1. Construct a plot.

 **\* Arranging** the data in order is not essential but

 recommended.

 **\* The** *leaves in the final stem and leaf plot* ***should be***

 *arranged in order.*

**Paired data and scatter plots:** A scatter plot is a graph of ordered pairs of data values that are used to determine if a relationship exists between the two variables.

**Analyzing the scatter plot:**

1. A positive linear relationship
2. A negative linear relationship
3. A nonlinear relationship
4. No relationship