Chapters 1.5 and 2.5

Statistics: Collecting and Displaying Data

Chapter Objectives:

- Decide which data is relevant, know methods to collect and organize data ;
- Design and use a data collection sheet or a questionnaire for a simple survey;
- Know the qualities of a good questionnaire;
- Know the difference between primary data and secondary data;
- Know the difference between quantitative data and qualitative data;
- Know the difference between discrete data and continuous data;
- Construct and use tally charts for individual data sets, frequency tables for grouped data (grouped appropriately) and two ways tables;
- Draw and be able to read:
    - Bar line graphs and histograms;
    - Pie charts;
    - Simple line graphs for time series;
    - Stem and leaf diagrams

I. Generalities. Types of data:

Definition: Statistics is the science of collecting, organizing and analyzing data.

Data is a collection of facts or pieces of information.

Primary data is data which you collect yourself.

Secondary data is data collected by someone else on your behalf, for example information that you find on internet or in articles.

Data can also be **quantitative** (in form of numbers), or **qualitative**, usually represented by adjectives (such as colors or attributes). In our class we will deal mainly with quantitative data.

Quantitative data can be: discrete and continuous.

Discrete quantitative data takes only specific values. For example, a survey to find the number of students in different classes can produce only discrete specific values.

In contrast, continuous data can take any value (usually in a specific range). For example, the height of a group of individuals or the timing of an event are examples of continuous data.

Example:

Do Exercise 5.1 in Book2.

II. Collecting data:

There are two main ways which we will investigate when collecting data:

the interview and the questionnaire.

In an interview, the interviewer asks questions and writes down the given answer. The questions do not have to be pre-determined before the interview. A questionnaire is a specific list of questions which people answer, usually on their own, without any other help.

Look at the advantages and disadvantages of interviews and questionnaires on Page 41 (Book 2) and summarize these in a table.

A questionnaire needs to follow the following ***basic rules***:

- Make it clear what you are trying to find out. For example: "Excuse me, I am conducting a survey to find out if people would like to see new traffic lights at the crossroads. Please would you help me by answering these simple questions?" ;
- Make the questions simple and give a choice for the answer.
- Do not have leading or biased questions;
- Do not ask sensitive or embarrassing questions.

Exercise:

Do Examples 5.1 A and 5.1B in Book 1 .

III. Organizing data:

Collected data can be organized and displayed in different types of graphs. In this class we will look at the following types of graphs / charts: Tally Charts, Pictograms, Grouped frequency charts (for grouped data), Bar Graphs, Histograms, Two Ways Tables, Pie Charts, Stem and Leaf Diagrams and Line Graphs. Let us briefly present each of these:

III.1 Tally Charts:

A tally chart works well for individual data. It records the number of results for each data item.

For example, a survey is carried out to test a manufacturer's claim that there are "about 36 chocolate pieces in each pack". The number of pieces in each pack (for 25 packs) is counted and the results are shown below:

35	36	34	37	36	36	38	37	36	35	38
34	35	36	36	34	37	38	37	36	35	36
36	37	36								

These results are displayed better in a tally chart shown below:

Number	Tally	Frequency
34	III	3
35	IIII	4
36	𝍸 𝍸	10
37	𝍸	5
38	III	3

The tally column is filled in as the survey is being carried out. The frequency column is completed by counting up the tally marks at the end of the survey.

Figure 1: Tally chart for the data above

Looking at the tally chart, we can see clearly the frequency (of number of pieces per pack) and validate the manufacturer's claim.

Is the manufacturer's claim true?

III.2 Pictograms:

Pictograms are very similar with tally charts, only that they usually use a symbol to display a certain number of items. This may help read the chart easier.

For the same data shown on Page 2, a pictogram is shown below:

Number of chocolate buttons	Frequency
34	◔
35	●
36	● ● ◖
37	● ◢
38	◔

Key
● = 4 chocolates

Look at the key to see what each symbol represents.

Figure 2: Pictogram for the data on page 2.

III.3 Grouped frequency table:

If there is a big range of the data, it is sometimes useful to group the data into a grouped frequency table.

For example, the ages of 30 residents in a care home are shown below:

98	71	76	77	72	78	77	73	76	86
75	79	81	105	100	74	82	88	91	96
85	90	97	102	83	101	83	84	80	95

This data can be displayed in a grouped frequency table, as shown below:

Age	Tally	Frequency
71–80	𝍸 𝍸 ||	12
81–90	𝍸 ||||	9
91–100	𝍸 |	6
101–110	|||	3

Figure 3: Grouped frequency table for the data above

The ages could also have been grouped in 71-75, 76 – 80, 81 – 85 ,etc. The decided size of each group is up to the collector, but it is important that the groups are of the same size, they do not overlap and that they complement each other.

III. 4 Bar Graphs and Bar – Line Graphs:

These are displaying the frequency of individual data items (or grouped data items) on vertical, as bars (or lines).

A bar graph of the frequency table shown in Figure 3 is displayed in Figure 4 below:

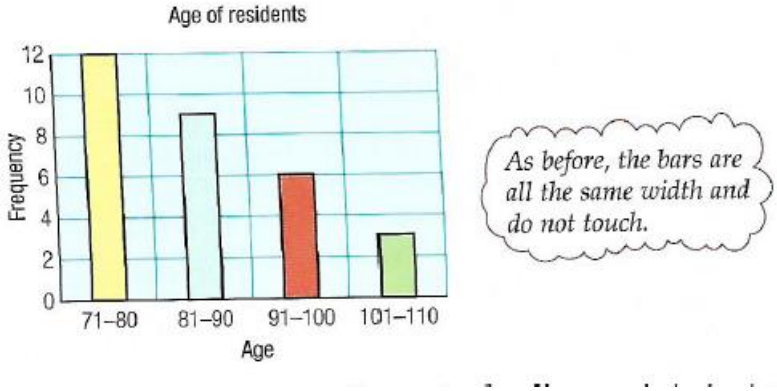

Figure 4: Bar graph of the frequency table shown in Figure 3

A bar – line graph for a different example of individual data items is shown on Page 46 of Book 1.

Examples:

From Exercise Set 5.2 in Book 1 do exercises 1, 2, 3, 4 and 5.

III.5 Histograms:

Histograms are very similar with bar graphs when displaying grouped data or individual data.

The main differences between histograms and bar graphs are:

- For histograms the bars are joined together;
- For histograms the bars can have varying width;
- For histograms the frequency of the data is represented by the area of each bar (with * height) and not by the height only. However, for most histograms , the bars are of equal width, and therefore the frequency is actually represented by the height of each bar, as it was for bar graphs as well.

Consider for example the results of 32 students on a Math test (scored out of 100 points) shown in the frequency table below:

Score	Frequency
1–10	0
11–20	0
21–30	1
31–40	2
41–50	5
51–60	8
61–70	7
71–80	6
81–90	2
91–100	1

A histogram of this grouped data is shown in Figure 5:

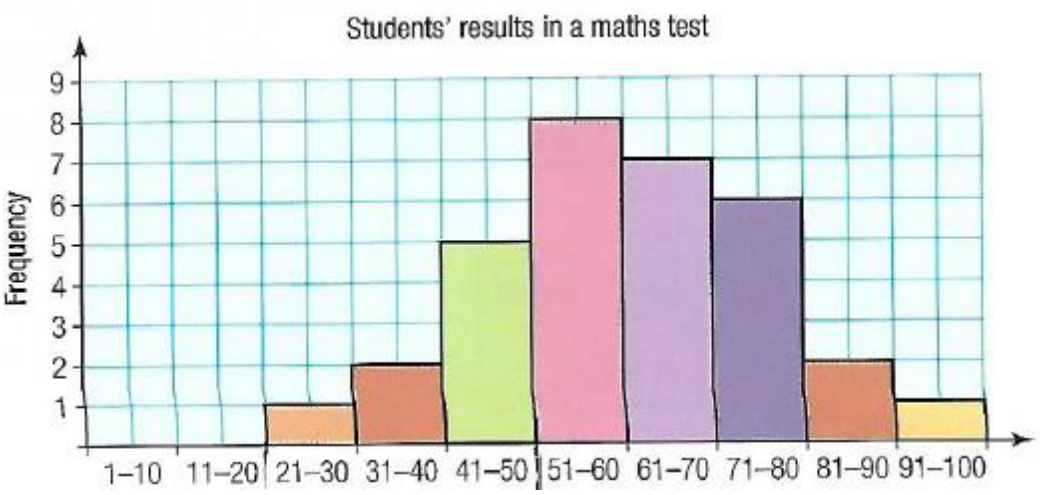

Figure 5: Histogram of the data on Page 5.

Examples:

From exercise set 5.3 (Book 2) do exercises 1, 2, 4 and 7. Do exercise 9 as bonus.

III.6 Two ways tables:

If data is split into two main categories (for example boys and girls, young and elderly, metal and non-metal etc), then two ways tables is the most appropriate way to display this data. For example, the two ways table in Figure 6 shows the numbers of boys and girls in each of Years 7-11 in a school:

		Year				
		7	8	9	10	11
Gender	Girls	25	31	19	34	37
	Boys	20	32	23	30	40

Figure 6: A two – ways table

Looking at the table above, answer:

- How many girls are in year 8 ?
- How many students are in year 10?
- How many boys are in total in years 7 -11 ?

Then do exercise 5.3 A from Book 2.

III.7 Stem and leaf diagrams:

Somewhat similar to two ways tables are stem and leaf diagrams. In this case, the "bars" or stems are made from the data itself and the "leafs" are individual data points for that category (for that stem).

For example, the data below records the heights (in cm) of 30 adults.

181	178	192	188	186	179	175	168	176	183
188	190	180	177	165	161	172	175	176	184
182	174	173	168	168	172	179	180	181	175

We display this information in a stem and leaf diagram shown in Figure 7:

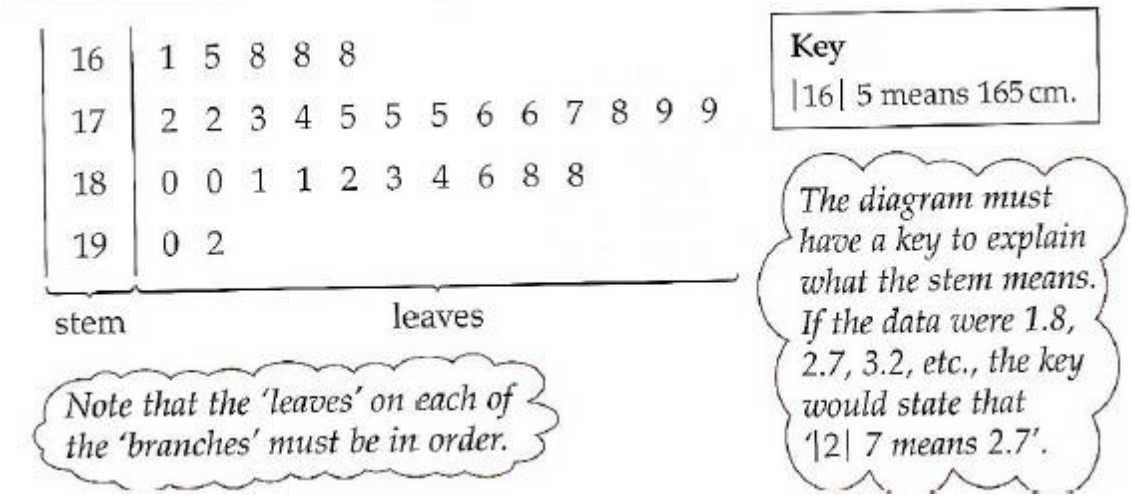

Figure 7: Stem and leaf diagram for the data above.

Examples:

From Exercise set 5.3 D (Book 2) do exercise 1 and 2.

III.8 Line graphs:

When data is collected over a period of time, we typically plot a "line graph" of this data, with the x axis representing time.

For example, the temperature (in degrees C) at a ski resort is recorded every 4 hours, and this data is shown below for one particular day:

Time	00.00	04.00	08.00	12.00	16.00	20.00	24.00
Temperature (°C)	–8	–11	–6	3	4	1	–5

Let us display this data as a line graph. The line graph is shown in Figure 8 below:

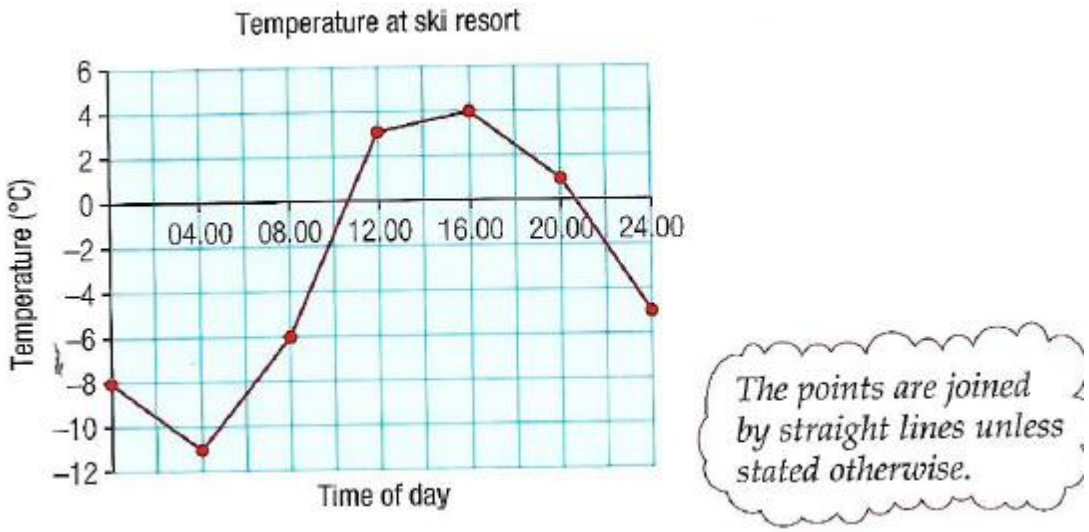

Figure 8: Line graph of the temperatures shown on Page 7.

Using this line graph above, what do you think that the temperature at 10.00 a.m. was?

Examples:

From Exercise set 5.3 E in Book 2, do exercises 1 and 3.

III.9 Pie Charts:

Pie charts are popular ways to represent data. In a pie chart, each sector (slice) represents a fraction of the total. This fraction is proportional to the frequency of that item (as a fraction of the total).

For example, the table below represents the different flavors of ice cream sold, and their relative fraction of the total (the relative frequency):

Flavour	Frequency	Fraction of the total
Vanilla	25	$\frac{25}{100}=\frac{1}{4}$
Chocolate	50	$\frac{50}{100}=\frac{1}{2}$
Strawberry	25	$\frac{25}{100}=\frac{1}{4}$

A pie chart for this data is shown in Figure 9 on the next page:

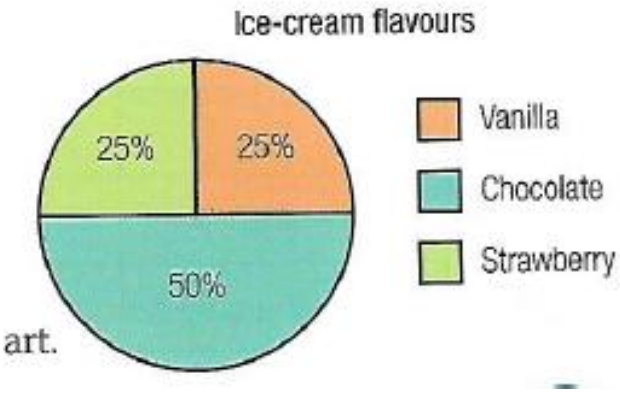

Figure 9: Pie chart of the data on page 8.

Therefore, to build a pie chart, we need to first calculate the relative frequency of each data (or grouped data) item, and then represent that as the particular central angle in a pie chart (calculate this angle as that particular fraction of 360^o).

Look at worked example on page 48 in Book 2 for a more involved example of a pie chart.

Examples:

From exercise set 5.3 C (Book 2) do exercises 1, 2 and 3 and from Exercise set 5.2 (Book 1) do exercises 6 and 7.

Other examples:

- For more examples of data to be displayed in pie charts, histograms, bar charts and so on, see: https://www.ck12.org/book/Basic-Probability-and-Statistics-A-Full-Course/section/7.2/
- See also problem 8 in Past Paper 1 (2013) : stem and leaf graph and problems 4, 8 and 19 in Past Paper 2 (2013)