

PROBABILIY AND STATISTICS I

LECTURE FOUR

Data presentation

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on the common methods of data presentation.

Intended learning outcomes

At the end of this lecture, you will be able to construct and interpret charts and graphs for different data types.

References

These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture

Data presentation

Statistical data is presented through pictograms and graphs. These are visual illustrations which reduce the complexity of the data.

- Pictograms:

This is convenient for a small data set. Data is represented using a relevant pictorial symbols, or neutral shapes.

- Graphs:

Some of the common graphs used in research include:

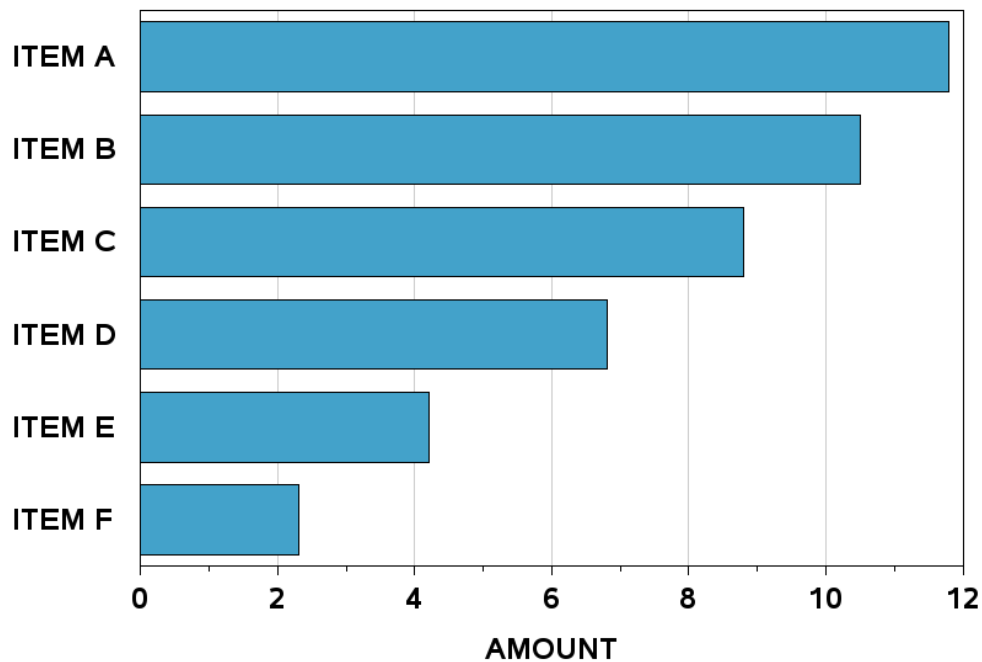
- **Bar graphs**

They are best used to represent nominal and ordinal data sets on a Cartesian plane having only one major axis giving statistical information, with the other remaining for identification purposes only. The height of each bar represents the number of items in

the given category. The bars can be vertical or horizontal. The width of the bars and the gap between the bars is always uniform throughout. Bar graphs can be of the following types

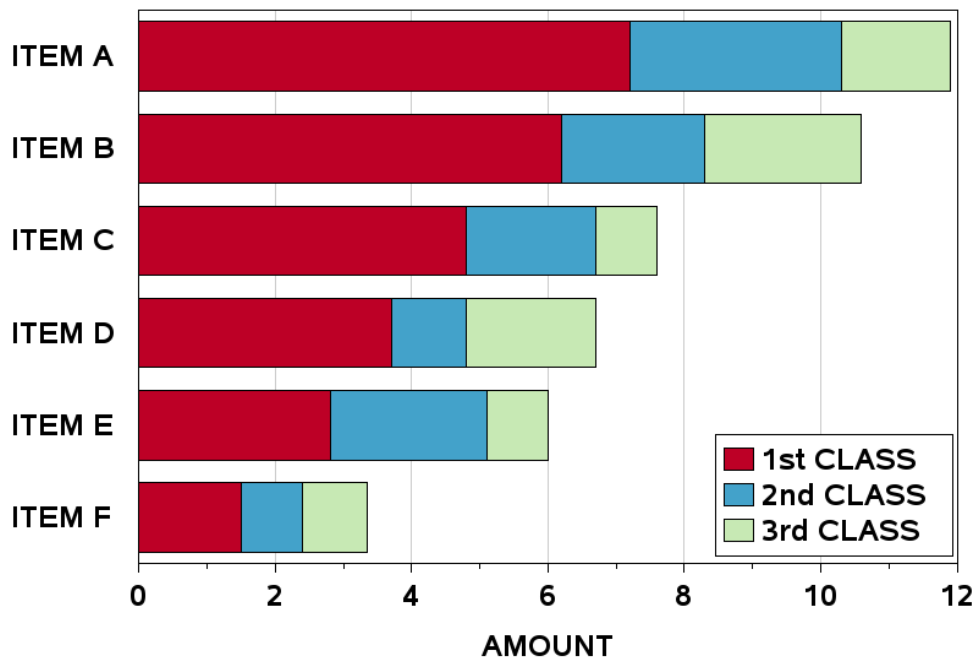
Simple bar graphs: - Used to represent only one category of data where only the lengths of the bars vary. The bars can either be vertical or horizontal

Simple Bar Chart

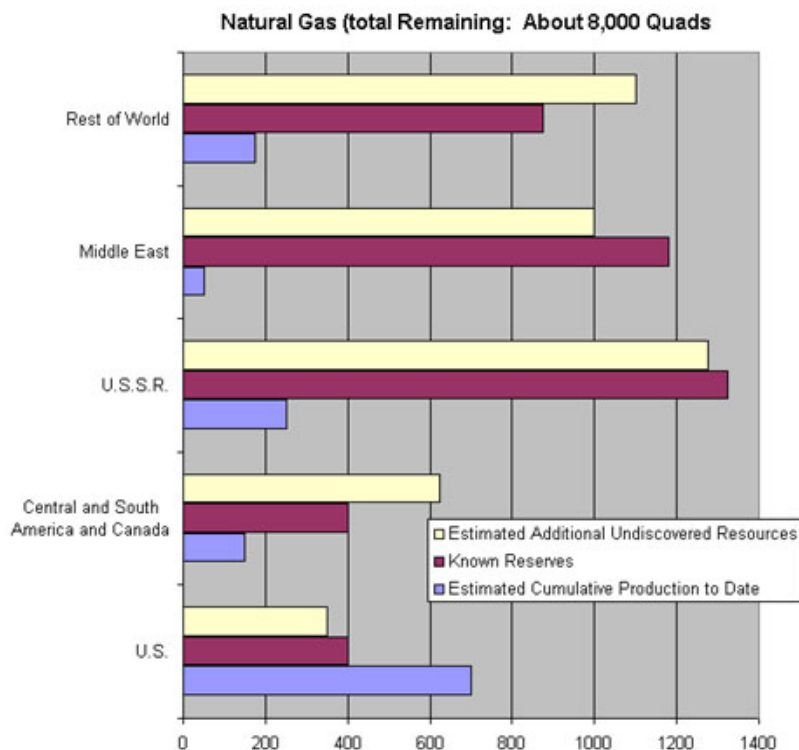


Subdivided bar graphs: - these are used to present distribution ratios diagrammatically. The bars are divided into more than one component. A key is then provided to distinguish the different components.

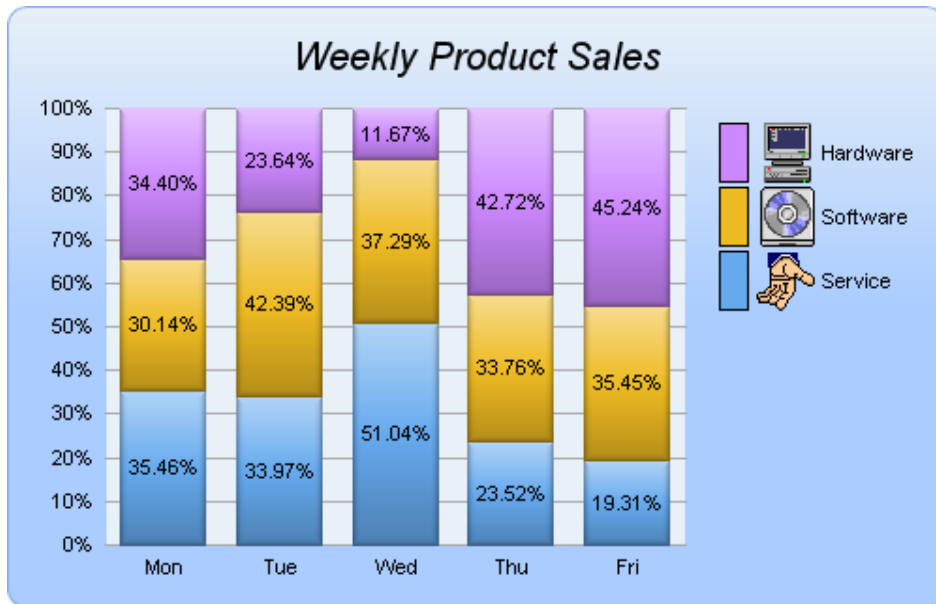
Subdivided Bar Chart



Multiple Bar graphs: - This is used to represent two or more interrelated data sets so as to allow for comparison between the variables. Different shades or patterns can be used to distinguish between the bars.



Percentage bars: - These are useful where there is need to portray the relative changes in data. The length of the bars is kept equal to 100 and segments are cut in the bars to represent the components (percentages) of an aggregate.



Deviation bars: - They are used to represent net quantities. The bars can have both negative (below the base line) and positive values (above the base line).

➤ Pie chart

This is the presentation of data in sectors with the sector angle being proportional to the amount of data. They are recommended for use when presenting less than five categories of data. The sector angle is obtained as:

$$\theta^{\circ} = \frac{x}{n} \times 360^{\circ}$$

Where θ° – sector angle; x – the number of items in a given category; n – the total number of data items.

➤ Histograms

These are special types of bar graphs that displays data by using contiguous vertical bars. It is a graph of class frequencies against lower class boundaries. The area of the each bar is proportional to the frequency represented. They are most suitable for interval and ratio scale data. The scale along the horizontal axis is continuous. The width of the bars is equal to the class interval.

If the class intervals are equal, the frequency is taken on the vertical axis. When the class intervals are unequal, then the vertical axis has the frequency densities. The frequency density is obtained as:

$$\text{Frequency density} = \frac{\text{frequency}}{\text{class interval}}$$

Histograms cannot be drawn for distributions having open ended classes.

Note: the distinction between a bar graph and a histogram is that a bar graph is one dimensional where only the height of the bars matter while a histogram is two dimensional, both height and width of the bars matter.

➤ Frequency polygon

This graph displays data using lines that connecting points plotted for the frequencies at the midpoints of the classes. Two or more frequency distributions can be compared on the same graph which cannot be done with histograms.

There are two types of frequency polygons:

- Simple frequency polygon – a graph of frequencies against the class midpoints
- Superimposed frequency polygon – this is obtained when the midpoints on top of the bars of a histogram are joined with straight lines.

Both ends of the polygon are closed by extending the line to the horizontal axis. This is done by including two hypothetical classes, one at each end, and each with a frequency of zero. The extended line is drawn to the midpoint of the two classes.

A smoothed frequency polygon can be drawn by joining the points freehand such that the area under the curve is approximately equal to the area under the polygon. It is

necessary to first draw the polygon then smooth it out. Smoothing is done to eliminate as far as possible all accidental variations that may be present in the data.

➤ **Cumulative frequency curve (Ogive)**

This is a graph of the cumulative frequencies for the classes in a frequency distribution against the upper class boundaries.

There are two methods for constructing the ogive

1. The less than method (more commonly used) - This starts with upper limits of the classes and go on adding frequencies.

For the data set

CAT Marks	No of students
10-14	2
15-19	12
20-24	6
25-29	3
30-34	1

We have the following using the less than method

Marks less than	No of students
14	2
19	14
24	20
29	23
34	24

When these cumulative frequencies are plotted against the upper class boundaries, we get a rising curve

2. The more than method - This starts with the lower limits of the classes and from the total frequencies we subtract the frequency of each class.

When the cumulative frequency curve is plotted, we get a declining curve.

Marks more than	No of students
10	24
15	22
20	10
25	4
30	1

Ogives are used to portray the number or proportion of cases above or below a given value. They are also used to obtain graphically values such as the median, quartiles, deciles etc.

Bibliography

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.