

Graphical Presentation for Statistical Data (Relevant to AAT Examination Paper 4: Business Economics and Financial Mathematics)

Y O Lam, SCOPE, City University of Hong Kong

Introduction

The most convenient and popular way of describing data is using graphical presentation. It is easier to understand and interpret data when they are presented graphically than using words or a frequency table. A graph can present data in a simple and clear way. Also it can illustrate the important aspects of the data. This leads to better analysis and presentation of the data. In this article, we discuss the approach for the most commonly used graphical methods such as bar charts, histograms, frequency polygons and cumulative frequency polygons.

Bar charts

Bar charts are used when comparing the values of multiple variables. Bars charts are presented using vertical or horizontal bars. Bars may be drawn separately from each other. It is important that the width of each bar should be the same to avoid misleading information.

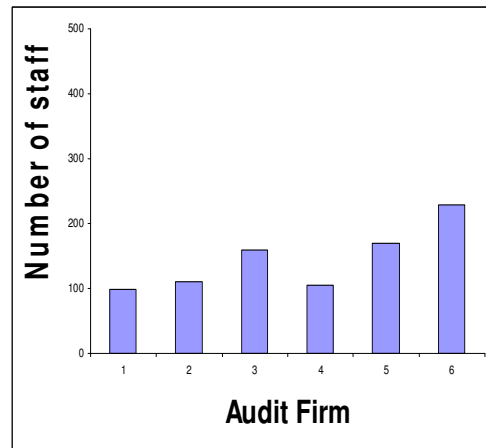
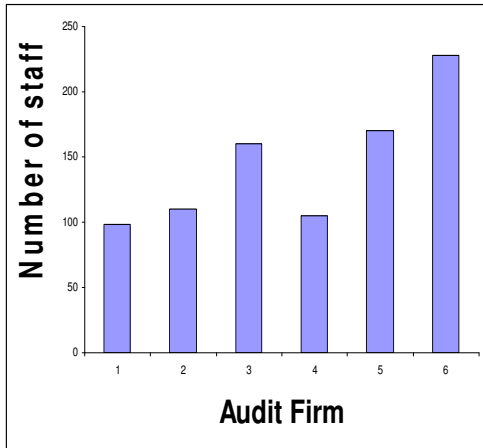
Figure 1 is an example of simple bar charts which mislead users by the manipulation of horizontal and vertical axes; this can affect the interpretations. **Simple bar charts** have two purposes. 1) They reflect the actual magnitude of the frequency of each item. 2) Their frequencies can be compared by comparing the heights of bars on the chart.

In Figure 2, **multiple bar charts** show the number of male and female staff in six audit firms. It is apparent that there are more female staff in the six audit firms.

Component bar charts are used to sum each component. Figure 3 is an example of a component bar chart. From this figure, we can see that total number of staff changes in different audit firms and with the components of each audit firm.

Component bar charts and multiple bar charts are similar, except for the following:

1. Component bar charts can show the grand total (in the above example, the total number of staff in each audit firm), whereas multiple bar charts cannot.
2. Component bar charts cannot show clear comparative magnitudes of each component, whereas multiple bar charts can.



i.

ii.

Figure 1. Simple bar chart. The two bar charts above represent the same data but they produce different interpretations by varying scales. (i) The maximum scale 250 is used and it is only slightly higher than the highest value in the study data. (ii) The maximum scale 500 is used and it is largely higher than the highest value.

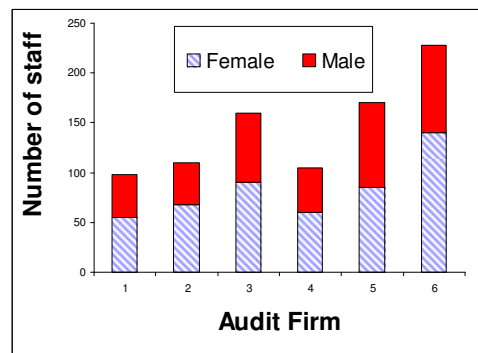
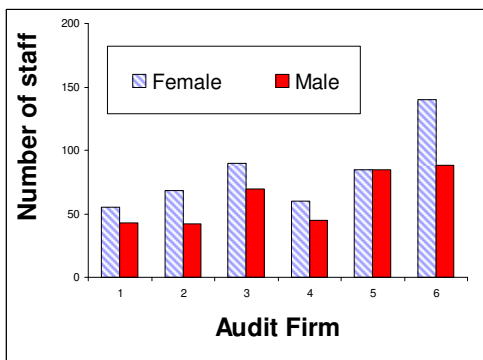


Figure 2. This *multiple bar chart* shows the number of male and female staff in six audit firms

Figure 3. This *component bar chart* shows the total number of staff in six audit firms

Histograms and frequency polygons

Histograms or column bar charts are common ways of presenting frequency in a number of categories. Commonly used graphical presentation methods also include the frequency polygon and ogive. Histograms portray an unequal width frequency distribution table for further statistical use. A frequency distribution table is a tabulation of the n measurements into mutually exclusive k classes showing the number of observations in each. The bars appear in a histogram where the classes are marked on the x axis and the class frequencies on the y axis. It is important to note that a bar chart does not have x -axis units. The histogram is constructed by creating x -axis units of equal size and these should correspond to the frequency table. Example 1 shows how to construct a frequency table.

Example 1

The following data have been obtained for the number of customers arriving per hour in a sample of 30 supermarkets. Construct a frequency table with class interval starting from 21 but less than 31, 31 but less than 41, ..., up to 71 but less than 81.

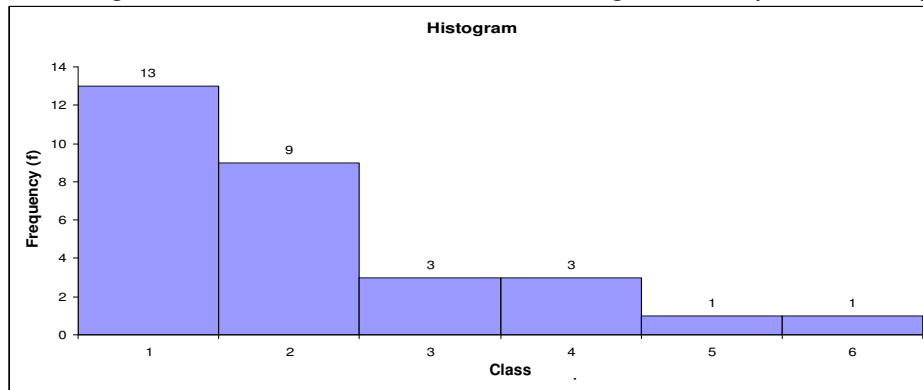
52	29	26	29	32	26
62	49	40	25	23	31
44	22	27	38	73	32
27	32	28	30	57	34
39	37	28	24	51	50

The first step is to count how often each value occurs in given class intervals. For example, 13 supermarkets might be in the class 21 but less than 31 customers per hour, 9 supermarkets in the class 31 but less than 41.... and so on. The cumulative frequency is calculated by adding the frequency for each class. An example of a frequency table, including both frequencies and cumulative frequency, is shown in Table 1

Table 1. Frequencies of the number of customers arriving at the supermarkets per hour

Class	Number of customers arriving per hour	Frequency (f)
1	21 but less than 31	13
2	31 but less than 41	9
3	41 but less than 51	3
4	51 but less than 61	3
5	61 but less than 71	1
6	71 but less than 81	1
		$\Sigma f = 30$

Figure 4. Histogram of number of customers arriving at the supermarkets per hour



We can see from the above histogram that the classes are marked on the x -axis and the class frequencies on the y -axis. The class frequencies are represented by the heights of the bars, and the bars are drawn adjacent to each other. Thus, a histogram describes a frequency distribution using a series of adjacent bars.

Based on the histogram in Figure 4, we can conclude that:

1. The lowest number of customers arriving per hour at the supermarkets is 21, and the largest is 81.
2. The class with the highest frequency is 21 up to 30 customers arriving per hour. A total of 13 of the 30 observations fall within this range.
3. 22 of the 30 supermarkets, or 73.3 percent, saw 21 to 40 customers arriving per hour.

Thus, frequency distribution can be easily interpreted using a histogram. We can reach the same conclusions with a frequency distribution table. If unequal class intervals are used in a frequency distribution, then the height of each bar is proportional to its frequency. This means that we use a *relative frequency* distribution instead of the actual frequencies. For example, if one of the class intervals is twice as wide as the other class intervals, the frequency of this class interval must be divided by two to become the actual height of the bar.

Histogram and simple bar charts are similar, except for the following:

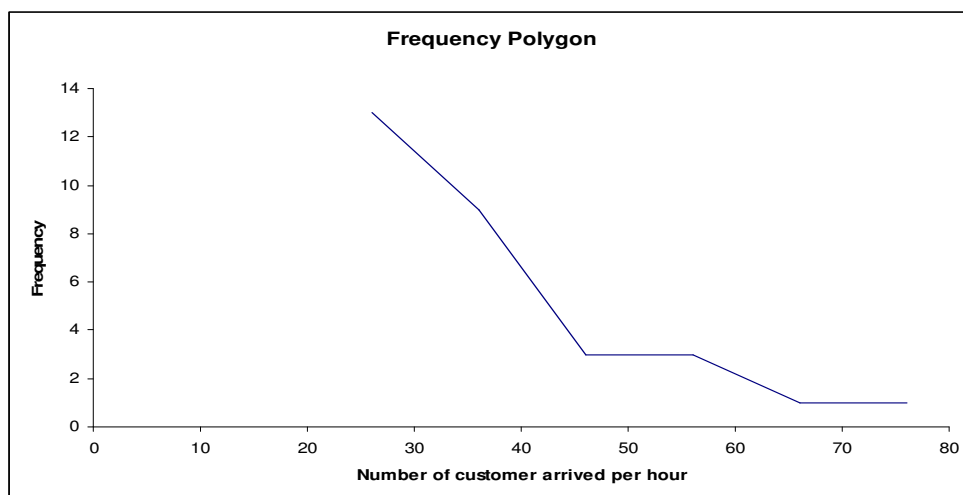
1. A histogram depicts a frequency distribution using a series of adjacent rectangles.
2. The height of each bar in a histogram is proportional to its frequency.
3. A simple bar chart can be used to show nominal and ordinal data, while a histogram is only used for interval-ratio data.
4. A histogram draws rectangles with class intervals on the x -axis and the corresponding frequencies on the y -axis.

Frequency polygons are not that commonly used when compared with histograms. These consist of line segments which are drawn by connecting the points formed by the intersections of class midpoints and class frequencies. The construction of a frequency polygon is illustrated in Figure 5.

Table 2. Midpoints and frequencies of the customers arriving at the supermarkets per hour

Class	Number of customer arriving per hour	Midpoint	Frequency(f)
1	21 but less than 31	26	13
2	31 but less than 41	36	9
3	41 but less than 51	46	3
4	51 but less than 61	56	3
5	61 but less than 71	66	1
6	71 but less than 81	76	1
			$\Sigma f = 30$

Figure 5. Frequency polygon of the customers arriving at the supermarkets per hour



Both histogram and the frequency polygon can provide a quick picture of the main characteristics of the study data. The frequency polygon has advantages over the histogram. First, it allows us to read the trend of the distribution. It is clear that the highest frequency is for 21 up to 30 customers arriving per hour and the frequencies fall as the number of customers arriving increases. Second, a frequency polygon is ideal for comparing two or more frequency distributions.

Ogive or cumulative frequency polygon

Lastly, if the readers are interested in knowing the number of observations “below” or “above” a certain value in the study data, this value can be estimated using a cumulative frequency distribution. The cumulative frequency distribution can be portrayed graphically in an ogive or cumulative frequency polygon.

To construct an ogive or cumulative frequency polygon, the first step is to calculate the cumulative frequencies. Referring to the frequency table, we can see that in 13 supermarkets less than 31 customers arrived per hour. Those 13 customers, plus the 9 in the next class, equal 22. This process continues for all the classes. To plot a

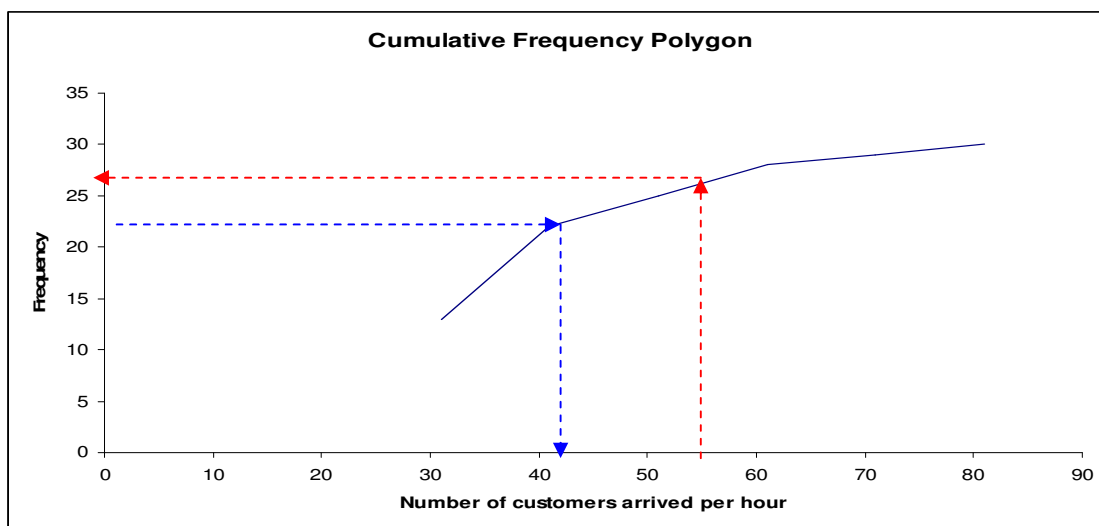
cumulative frequency polygon, scale the upper limit of the class along the x-axis and the cumulative frequencies along the y-axis.

Table 3. Cumulative frequencies of the customers arriving at the supermarkets per hour

Class	Number of customers arriving per hour	Frequency (f)	Cumulative frequency	Workings
1	21 but less than 31	13	13	
2	31 but less than 41	9	22	=13 + 9
3	41 but less than 51	3	25	=13 + 9 + 3
4	51 but less than 61	3	28	=13 + 9 + 3 + 3
5	61 but less than 71	1	29	
6	71 but less than 81	1	30	

$\Sigma f = 30$

Figure 6. Cumulative frequency polygon of the customers arriving at the supermarkets per hour



To find the value below a frequency of 22, a horizontal line is drawn from the value of 22 to the polygon, and then dropped down to the x-axis and the value read. It is about 41, so 22 supermarkets have less than 41 customers arriving per hour. The percentage of supermarkets for which less than a specified number of customers arrived can also be estimated. Suppose we need to know the number of supermarkets which had less than 55 new customers per hour. First, locate the value 55 on the x-axis: draw a vertical line from this to the polygon, and then a horizontal line to the vertical axis on the left. The value is about 26, so 26 supermarkets have less than 55 new customers per hour.