

## Chapter 2

# Presentation of Statistical Data

**Abstract** Data are collected often in raw form. These are then not useable unless summarized. The techniques of presentation in tabular and graphical forms are introduced. Some illustrations provided are real-world examples. Graphical presentations cover bar chart, pie chart, histogram, frequency polygon, pareto chart, frequency curve and line diagram.

**Keywords** Presentation • Table presentation • Graph presentation • Types of presentation

Data are often collected in raw form. These are then not useable unless summarized. There are certain guidelines for data summarization such as summarization

- should be as useful as possible,
- should represent data fairly, and
- should be easy to interpret.

After collection of data (primary or secondary), it is necessary to summarize them suitably and present in such forms as can facilitate subsequent analysis and interpretation. There are two major tools/techniques for presentation of data as follows:

- Presentation in tabular form
- Presentation in graphical form.

## 2.1 Tabular Presentation

Data may be presented in the form of statistical tables. In one table only simple frequencies can be shown. Also, in the same table cumulative frequencies, relative frequencies, and cumulative relative frequencies can be shown. Relative frequencies and cumulative frequencies are defined as follows:

Relative frequency: It means the ratio of the frequency in the category of concern to the total frequency in the reference set.

$$\begin{aligned}\text{Relative frequency of } X_i &= \frac{\text{actual frequency of } X_i}{\text{sum of all frequencies}} \\ &= \text{proportion}\end{aligned}$$

Relative frequency of  $X_i$  (%) = proportion \* 100

Cumulative frequency of  $X_i$  = sum of all frequencies of all values up to and including  $X_i$ .

In the same table the simple frequencies combined with one or more but not all, of the cumulative frequencies, relative frequencies, and cumulative relative frequencies may be shown. Table 2.1 serves as an example of tabular presentation of simple frequency distribution. In the same table all have been shown together for illustrative purposes.

In Table 2.1 only one dimension has been shown. But intelligently more than one dimension may also be shown in the same table. This is demonstrated in Table 2.2.

There are advantages in such table presentation. One advantage is that more than one parameter (here education and occupation) can be shown in the same table. This serves as a concise presentation. Another advantage is easy comparison and interpretation. In Table 2.2 both fathers' and mothers' situations against each of the variables of the parameters can be readily compared and interpreted.

In the examples provided in Tables 2.1 and 2.2, frequencies have been shown against each categories. The categories served as groups, which were predetermined on the basis of certain criteria. More specifically, these are presentations of categorical data. But in practice it often becomes necessary to group the metric data to form some groups or categories or classes. Here the point of interest is to see how well the data can be grouped or classed. There are certain guidelines that help in grouping the data. The guidelines are

**Table 2.1** Types of organizations in which AIT alumni (1960–1987) are working

Type of organization	Freq.	Cumul. freq.	Relative freq.	Cumul relative freq.
Govt. office at central level	16	16	13.01	13.01
Govt. office at regional level	5	21	4.07	17.08
Public state enterprise	12	33	9.76	26.84
Private enterprise	32	65	26.02	52.86
Educational institution	31	96	25.20	78.06
Nongovt. organization	8	104	6.50	84.56
International organization	9	113	7.31	91.87
Others	10	123	8.13	100.00
Total	123		100.00	

**Table 2.2** Educational and occupational status of AIT alumni (1960–1987) parents

Status and levels	Father		Mother	
	<i>f</i>	%	<i>f</i>	%
<i>Education</i>				
Formal education	56	10.6	115	21.8
Primary education	105	19.9	182	34.5
Secondary education	119	22.5	112	21.2
Post secondary schooling	67	12.7	43	8.1
College education	70	13.2	38	7.2
University education	108	20.5	29	5.2
NA	3	0.6	9	1.7
Total	528	100.0	528	100.0
<i>Occupational</i>				
Farming	71	13.4	63	11.8
Commerce	119	22.5	97	18.4
Industry	16	3.0	9	1.7
Public service	207	39.2	45	8.5
Private service	75	14.2	36	6.8
Teaching	4	0.8	4	0.8
Multiple job	13	2.5	1	0.2
Housework	0	0	204	38.6
Other	15	2.8	9	1.7
NA	8	1.5	60	11.4
Total	528	100.0	528	100.0

Source AIT, AIT Alumni 1961–1987

- every score must fit into exactly one class,
- intervals should be nice, and
- classes should preferably be of the same width.

A frequency distribution is a more compact summary of data than the original observations. To construct a frequency distribution, we need to divide the range of the data into intervals known as classes. As already mentioned, the class intervals, whenever possible, should be of equal width, to enhance the visual information in the frequency distribution. We need to apply our judgment in selecting the number of classes in order to give a reasonable display. The number of classes used depends on the number of observations and the amount of dispersion in the data. Too few or too many classes are not very informative. It has been found that the number of classes between 5 and 20 is satisfactory in most of the cases. Also, the number of classes should increase with the number of observations. The number of classes may be chosen to be approximately equal to the square root of the number of observations. Thus, no. of classes =  $\sqrt{n}$  (approx.).

It is convenient to use a single nonoverlapping type of class for all types of data (discrete or continuous). Look at the following example (data in Table 2.3).

**Table 2.3** Crushing strength (psi) of bricks

215	147	296	230	215	150	171	215	211	228
155	236	267	192	204	185	126	212	198	200
213	224	192	210	231	196	198	221	210	215
257	193	208	271	244	278	213	195	224	220
170	181	226	178	173	246	181	251	287	248
218	217	250	200	210	226	284	230	200	207
210	231	158	249	258	214	250	224	228	160
184	215	137	208	185	219	215	203	204	230
249	164	214	217	233	185	222	237	224	219
165	268	221	243	227	240	233	208	225	201

No. of observations            100  
Approx. no. of classes     $\sqrt{100} = 10$   
The highest value            296  
The smallest value            126  
Range                             $296 - 126 = 170$

If the lowest and the highest values in the frequency distribution are chosen to be 120 and 300, respectively, the range becomes 180 (i.e.,  $300 - 120$ ). So, nine classes are chosen and consequently the class width is  $180/9 = 20$ . The results of this classification are shown in Table 2.4.

Another tabular presentation often used is cross tabulation. Cross tabulation is a joint frequency distribution of different values of two (or more) variables. Table 2.5 is an example of cross tabulation of two variables, namely “Academic Divisions” and “Levels of Satisfaction with Selected Aspects.” The figures in the table are indexes of satisfaction.

Very frequently, statistical test results are also presented in the form of tabular presentation. Table 2.6 is an example.

**Table 2.4** Crushing strength (psi) of bricks (classes)

Class interval	Frequency
$120 \leq x < 140$	2
$140 \leq x < 160$	4
$160 \leq x < 180$	7
$180 \leq x < 200$	13
$200 \leq x < 220$	32
$220 \leq x < 240$	24
$240 \leq x < 260$	11
$260 \leq x < 280$	4
$280 \leq x < 300$	3
Total	100

**Table 2.5** AIT alumni’s satisfaction with selected aspects of thesis research conducted at AIT

Academic division/period	Level of satisfaction with selected aspects		
	Flexibility in topic selection	Practical applicability	Support service
<i>Division</i>			
AFE	0.78	0.71	0.75
CA/CS	0.58	0.64	0.63
CRD/SE	0.57	0.52	0.50
EE	0.75	0.70	0.73
ET	0.76	0.76	0.74
GTE	0.76	0.70	0.67
HSD	0.76	0.71	0.66
IEM	0.75	0.63	0.65
SEC	0.66	0.64	0.69
WRE	0.67	0.67	0.70
<i>Period</i>			
Before 1965	0.58	0.73	0.55
1966–1970	0.58	0.56	0.60
1971–1975	0.70	0.63	0.65
1976–1980	0.72	0.67	0.67
1981–1985	0.74	0.69	0.72
1986–1997	0.72	0.71	0.70
Average	0.71	0.68	0.68

Source AIT, AIT Alumni 1961–1987

The main purpose of tabular presentation, in fact of all aspects of presentation, is that summary or presentation should be informative and meaningful. It should facilitate interpretation and subsequent analysis. To highlight this aspect, data in Table 2.7 are used.

During the several years from 1790 to 1984, both the urban and rural population of the United States maintained a steady increase. From the figures in column 4, it is also clear that the percentage of urban population continued to increase steadily. This indicates, together with the figures in columns 2 and 3 that the urban population grew at a faster rate compared to the rural population.

The underlying purpose in making the analysis of the data presented in Table 2.7 is to demonstrate that every table prepared should convey some interpretation message. Mere presentation of data in some tables is meaningless unless such purpose is served.

**Table 2.6** Attitudes toward necessity of education for men—results of multiple regression analysis

<i>X</i>	<i>B</i>	$\alpha(\beta)$	<i>T</i> -value	C.I. (%)
A	3.9745	0.0963	41.270	100.00
HH	−0.0171	0.0081	− 2.104	94.00
<i>Occupation</i>				
Agr	−0.0404	0.0491	− 0.822	59.00
Trd	−0.1678	0.0819	− 2.046	96.00
Srv	−0.0146	0.0518	− 0.282	22.00
Age years	0.0017	0.0014	1.202	77.00
Sch years	0.0089	0.0054	1.645	90.00
Test statistics				
$R^2 = 184$ df: Regression = 6				
<i>F</i> -value = 2.48 Residual = 66				
Confidence level = 97.00 %				

*Notes**X* = independent variables*a* = constant term $\beta$  = coefficient $\alpha(\beta)$  = standard error of  $\beta$ 

C.I. = confidence level

HH = household member

Agr = agriculture

Trd = trade

Srv = service

Rs = respondent

sch = schooling

**Table 2.7** Urban population growth in the United States

Year	Urban population (million)	Rural population (million)	Urban as a percentage of total population
1790	0.2	3.7	5
1810	0.5	6.7	7
1830	1.1	11.7	9
1850	3.5	19.6	15
1870	9.9	28.7	25
1890	22.1	40.8	35
1910	42.0	50.0	46
1930	69.0	53.8	56
1950	96.5	54.2	64
1970	149.8	53.9	73
1975	155.9	57.3	73
1980	170.5	56.1	74
1984	179.9	56.2	76

*Source* U.S. Bureau of the census. The census bureau as cities and other incorporated places, which have 2500 or more inhabitants define “Urban areas”

## 2.2 Graphical Presentation

Data presented in the form of tables give good information in concise form. Tables provide all relevant information of the data. Apart from tabular presentation, graphical presentation of data has also become quite popular. It gives visual information in addition to magnitudes. Furthermore, comparisons and changes in the data can be well visualized when presented in graphical form. A very useful part of graphical presentation is the interpretation of the graphs. In every graph we should try to interpret the data.

With the help of computer software packages such as Harvard Graphics, Lotus 123, Energraphics, etc., graphical presentation of data can be made in a variety of ways. But these may broadly be categorized into the following:

- Bar chart
- Pie chart
- Histogram
- Frequency polygon
- Pareto chart
- Frequency curve
- Line diagram.

### 2.2.1 Bar Charts

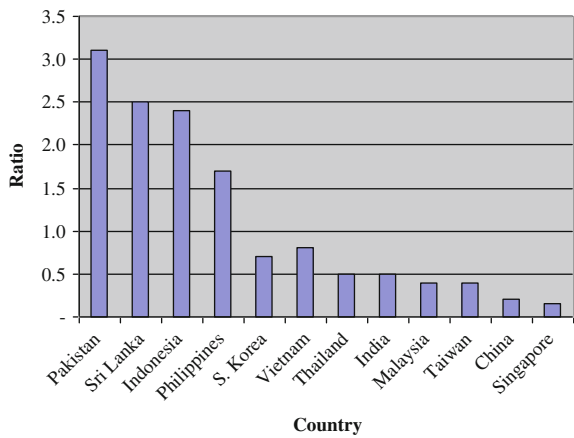
Bar charts are used for categorical data or metric data that are transformed into categorical data. Categories are shown on the horizontal axis. Frequency, percentage, or proportion is shown on the vertical axis. Bars are separated from each other to emphasize the distinctness of the categories. The bars must be of the same width. The length of each bar is proportional to the frequency, percentage, or proportion in the category. Levels ought to be provided on both axes.

In one figure only one variable can be depicted. This is illustrated in Figs. 2.1 and 2.2. Two or more variables can also be depicted in the same figure for ease of comparison. Figures 2.3 and 2.4 show presentation of two bars and Fig. 2.5 shows triple bar presentation.

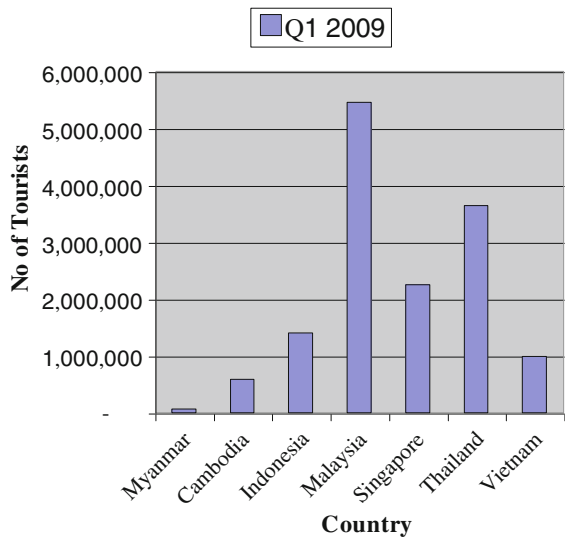
### 2.2.2 Pie Charts

Like bar charts, pie charts are also used for categorical data. A circle is divided into segments, the areas of which are proportional to the values in the question. But the areas are proportional to the angles the corresponding segments make at the center

**Fig. 2.1** Foreign debt/foreign reserve of some countries (2008)



**Fig. 2.2** Foreign tourist arrivals in Southeast Asia (Q1 2009)



of the circle. Thus, segments of the circle are cut in such a way that their values are proportional to the angles.

In one pie chart only values of one variable can be shown. However, two or more pie charts may be constructed side by side for comparison or to study the change over time. In Fig. 2.6 Thailand population (2009) is shown. Figure 2.7 is another example of pie chart presentation (Table 2.8).



**Table 2.8** Newly constructed dwellings in Bangkok and adjoining provinces

Type	1987			1991		
	Units	%	Angle	Units	%	Angle
Individual homes	34,679	65.0	234	35,604	27.6	99
Town house	16,326	30.6	110	52,116	40.4	146
Flat and condominium	1,707	3.2	12	39,861	30.9	111
Twin house	641	1.2	4	1,419	1.1	4
Total	53,353	100	360	129,000	100	360

### 2.2.3 Histogram

Histograms are used for metric data but converted to categories. These are somewhat similar to bar charts. However, there are some important features in histograms. The blocks in histograms are placed together one after another. These are not separated. Classes are ordered on the horizontal axis, with scores increasing from left to right. Areas of the blocks are proportional to the frequencies. If the class intervals are of equal width, the heights of the blocks/rectangles are proportional to the frequencies. If the class intervals are of unequal width, the blocks/rectangles are drawn in such a way that the areas of the blocks/rectangles are proportional to the frequencies. However, it is easier to interpret the histograms, if the class intervals are of equal width.

Data of Table 2.4 are used here to construct a histogram shown in Fig. 2.8.

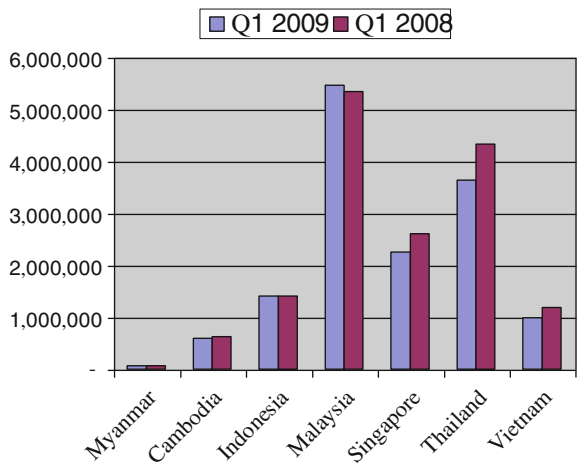
### 2.2.4 Frequency Polygon

It is also a graphical presentation of frequency distribution. It is more convenient than the histogram. The midpoints of the upper extremes of the blocks of the histogram are joined by straight lines. The first and the last parts of the polygon are to be brought to the horizontal axis at a distance equal to half of the class width.

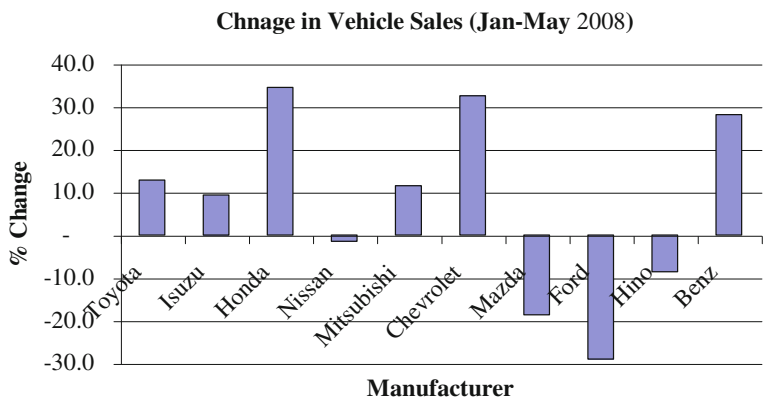
Data of Table 2.4 are used here for construction of the frequency polygon shown in Fig. 2.9.

### 2.2.5 Pareto Chart

A pareto chart is a bar chart for count (discrete) data. It displays the frequency of each count on the vertical axis and the count type on the horizontal axis. The count types are always arranged in descending order of frequency of occurrence. The most frequent occurring type is on the left, followed by the next-most frequently occurring type, and so on. Bars are placed side by side with no gap between the



**Fig. 2.3** Comparison of foreign tourist arrivals (Q1 2008 and Q1 2009). *Data Source* Pacific Asia Travel Association



**Fig. 2.4** Change in vehicles sales. *Data Source* Toyota Motors Thailand

adjacent ones. A segmented line is also drawn to depict the relative cumulative frequency distribution.

Pareto charts are useful, among other uses, in the analysis of defect data in manufacturing system, construction management, and others, and is an important part of quality improvement program since it allows the management and engineers to focus attention on the most critical defects in a production or process.

Data in Table 2.9 are used to construct the pareto chart shown in Fig. 2.10.

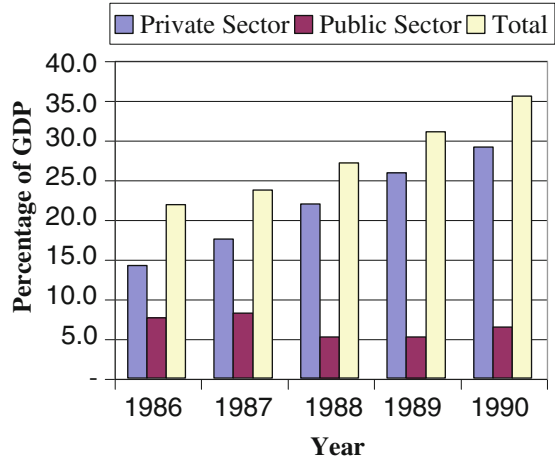


Fig. 2.5 Investment as a percentage of GDP. *Data Source* NESDB

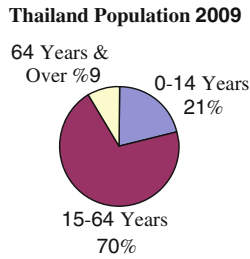


Fig. 2.6 Thailand population (2009). *Data Source* National Statistical Office

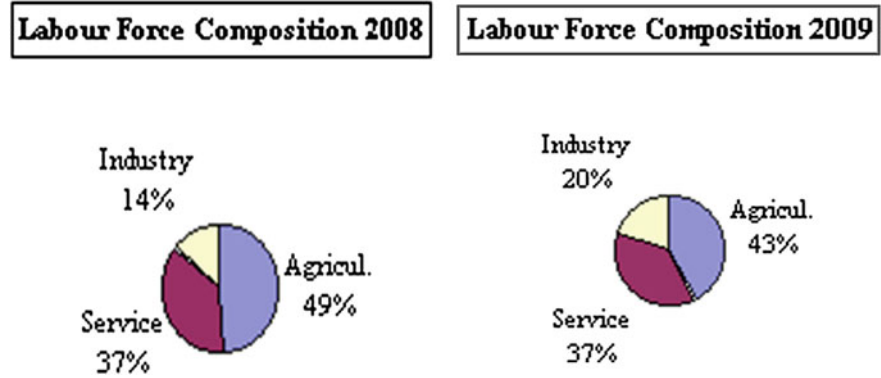
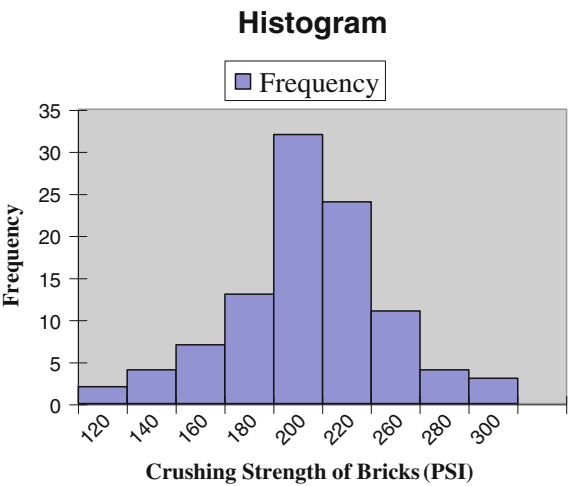
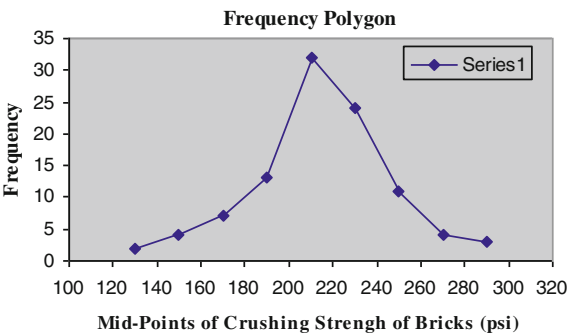


Fig. 2.7 Thai labor force by occupation. *Data Source* National Statistical Office

**Fig. 2.8** Crushing strength of bricks (histogram)



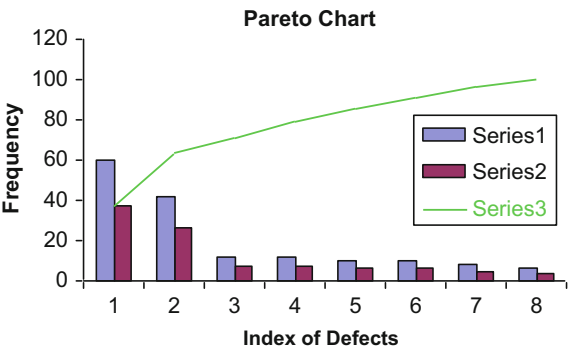
**Fig. 2.9** Crushing strength of bricks (frequency polygon)



**Table 2.9** Defects in building construction

Code defects	Frequency	Relative frequency (%)	R.C.F. (%)
Plaster	60	37.5	37.5
Curing	42	26.2	63.7
Flooring	12	7.5	71.2
Door	12	7.5	78.7
Distemper	10	6.3	85.0
Power line	10	6.3	91.3
Plumbing	8	5.0	96.3
Others	6	3.7	100.0
Total	160	100.0	

**Fig. 2.10** Defects in building construction (pareto chart)



### 2.2.6 Line Diagram

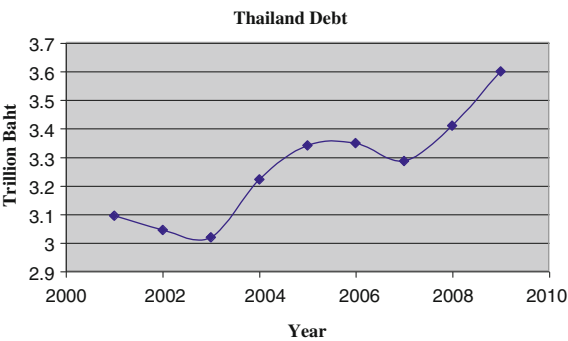
Line diagrams are drawn by plotting the values of two continuous variables. These show trends or changes in one variable resulting from changes in the other. One important application of the line diagram is to study the changes of various economic indicators over time. Line diagrams may be presented in the form of continuous lines or segmented lines depending on the phenomenon under study. Figures 2.11, 2.12 and 2.13 will serve as examples.

### 2.2.7 Frequency Curve

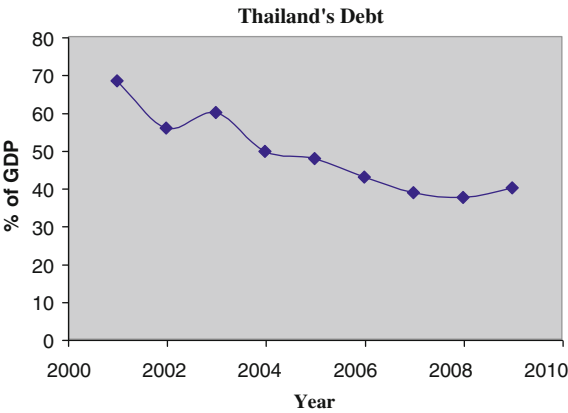
Frequency curve is a smoothed frequency polygon. It is produced by plotting the absolute frequency of an infinitesimally small range of a continuous variable. It is a theoretical distribution.

An example of frequency polygon is given in Fig. 2.14.

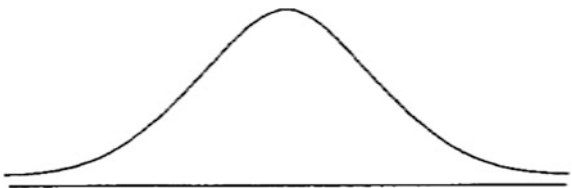
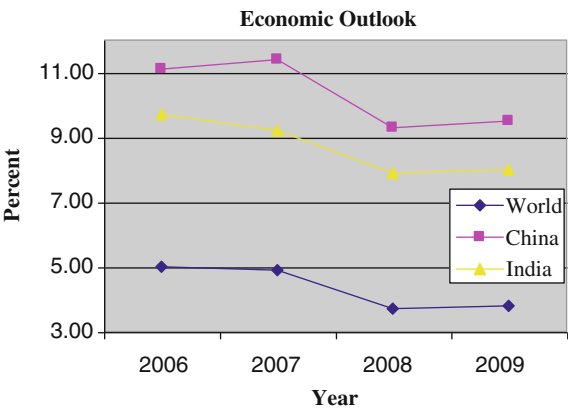
**Fig. 2.11** Thailand debt situation. *Data Source* Finance ministry



**Fig. 2.12** Thailand debt as percentage of GDP. *Data*  
*Source* Ministry of Finance



**Fig. 2.13** Comparison of economic outlook. *Data*  
*Source* International Monetary Fund



**Fig. 2.14** Frequency curve

## Problems

- 2.1 The Canadian International Development Agency (CIDA) has been providing financial support to AIT. The overall financial status as of 31 March 1991 is shown below.

Project activity	Planned expend	Actual expend	Actual/planned expend
Scholarship	4400	2270	51.6
Student research	387	108	27.9
Exploratory research	250	106	42.4
Demonstration project	250	58	23.2
Dissemination project	450	134	29.8
Seminar/workshop	400	75	18.8
Project support	150	59	39.3
Total	6287	2810	44.7

*Note:* The expenditure figures are in 1000 Canadian dollars

Construct bar charts for the expenditures as well as the proportions. Interpret the charts in comparing the planned and actual expenditures.

- 2.2 The number of enrolled students (as of May 1987) in the College of Medical Technology and Nursing, University of Tsukuba, Japan is as follows:

Level	Total no. of students	Female students
Undergraduate	7969	2194
Master degree program	1140	260
Doctoral degree program	1228	204
Laboratory school	4536	0
Medical technology	361	352
Total	15,234	3010

Construct pie charts to depict the distributions. Interpret the charts

- 2.3 The actual expenditures under CIDA activities in AIT during the four consecutive periods are shown:

Project component	Actual expenditures in 1000 CDN\$			
	2nd quarter 1990	3rd quarter 1990	4th quarter 1990	1st quarter 1991
AGP	82	74	94	162
EPM/NRP	133	88	24	99
HSD	148	169	45	100
Project support	6	6	7	5
Total	369	337	170	366

Draw segmented line diagrams for all the four project components in one chart. Interpret the results.

2.4 A quality control manager obtained samples to check the number of defective products. The number of defective products noted was as follows:

4	7	5	8	7	8
3	11	2	7	5	9
7	16	16	12	14	5
6	12	14	11	4	9
4	10	6	13	9	6
13	12	16	15	12	17
3	5	10	20	4	19
10	8	12	9	7	12
2	12	7	3	12	11
7	7	7	6	14	8
3	15	5	4	5	10
8	6	5	6	7	9

Group the data. Construct a histogram. Draw the frequency polygon. Comment on the distribution.

2.5 Electrical power demand in Dhaka was noted in two sample occasions—one in winter and the other in summer. The recorded demands were as follows:

Time (h)	Winter demand (MW)	Summer demand (MW)
16:00	972	1141
17:00	1203	1161
18:00	1519	1147
18:30	1551	1251
19:00	1549	1334
19:30	1475	1343
20:00	1475	1318
21:00	1317	1344
22:00	1170	1281
23:00	908	1227
24:00	841	1184

- (a) Draw smooth line charts to depict the trend of power demand over time.
- (b) Compare the two demands.



## References

- Asian Institute of Technology, *AIT Alumni 1961–1987 Tracer Study*, Bangkok: AIT, 1990, p.9
- International Monetary Fund
- Ministry of Finance, Royal Thai Government, Thailand
- NESDB, National Economic and Social Development Board, Thailand
- National Statistical Office, Thailand
- Pacific Asia Travel Association
- Toyota Motors, Thailand
- U.S. Bureau of the Census. The Census Bureau as cities and other incorporated places, which have 2500, or more inhabitants define “Urban areas”

Applied Statistics for Social and Management Sciences

Miah, A.Q.

2016, XV, 444 p. 73 illus. in color., Hardcover

ISBN: 978-981-10-0399-8