

PAGE MISSING (145 TO 152)

UNIVERSAL
LIBRARY

OU_156303

UNIVERSAL
LIBRARY

OUP-552-7-7-66-10,000

OSMANIA UNIVERSITY LIBRARY

Call No. 311

Accession No. P.G

Author 988. Curtis 15917

Title Statistics

This book should be returned on or before the date last marked below.

STATISTICS
AS APPLIED TO ACCOUNTING DATA

AUSTRALIA

**The Law Book Co. of Australasia Pty. Ltd.
Sydney : Melbourne : Brisbane**

CANADA AND U.S.A

**The Carswell Company Ltd.
Toronto**

INDIA

**N.M. Tripathi Ltd.
Bombay**

NEW ZEALAND

**Legal Publications Ltd.
Wellington .**

PAKISTAN

**Pakistan Law House
Karachi**

STATISTICS

AS APPLIED TO ACCOUNTING DATA

ADAPTED BY

C. R. CURTIS, M.Sc., Ph.D., F.C.I.S.

→
from

Advanced Accounting

By

R. KEITH YORSTON, F.C.A.(Aust.)

E. BRYAN SMYTH, A.I.C.A.

S. R. BROWN, F.C.A.(Aust.)

LONDON:

SWEET & MAXWELL LIMITED

2 and 3 Chancery Lane, London, W.C.2

1953

CONTENTS

<i>Chapter</i>	<i>Page</i>
1. PRELIMINARY CONSIDERATIONS	1
2. THE AVERAGE	24
3. DISPERSION	41
4. CORRELATION	57
5. TREND AND SEASONAL VARIATIONS	71
6. GRAPHIC PRESENTATION OF DATA	99
7. SAMPLING, ACCURACY AND APPROXIMATION, AND INDEX NUMBERS	134
<i>Appendix</i>	
A. LOGARITHM TABLES	171
B. ANNUITY TABLES—PRESENT VALUE OF ANNUITY OF £1 ..	152
C. ANNUITY TABLES—PERIODIC PAYMENT REQUIRED TO EX- TINGUISH £1 AND INTEREST	153

CHAPTER 1

PRELIMINARY CONSIDERATIONS

INTRODUCTION	1
Objects of Statistics.									
Value of Statistics.									
Limitations of Statistics.									
Internal Statistics.									
External Statistics.									
THE COLLECTION OF THE DATA	7
Market Research.									
THE ARRANGEMENT OF THE DATA	16
The Statistical Table.									
The Statistical Series.									
The Frequency Distribution.									

INTRODUCTION

To state that the operations of industry and commerce with their associated sphere, finance, are symbolically represented by figures, may indeed be to utter a truism, but the fact should nevertheless be emphasized in order to appreciate the close relationship between the increasing complexity of modern business on the one hand and the advances in technique required of the accountant on the other.

With the growth in size of financial structures, commercial adventures and industrial projects, with increased governmental controls and regulations respecting price-fixing, taxation, etc., it follows that accounting procedures must vary and expand. It is the accountant who gives effective expression to the wants of the community in matters of finance and as those wants increase there is a correspondingly increasing burden placed upon the accountant to arm himself with the technique necessary to deal with such needs.

Objects of Statistics

In order to keep abreast with such progress, the accountant has acquired from the field of statistics a fund of knowledge which assists him, not only in collecting his facts and in sorting, grouping and comparing them, but also, in analysing and interpreting the data in various ways depending upon the purpose in any particular case. These form the objects of statistics as applied to accounting.

Statistical work does not aim to eliminate altogether the human element in judgment. It does, however, by the very nature of its analytical method, tend to control it.

It is perhaps necessary to point out at this stage that the subject of statistics, both in its mathematical theory and practical application, is a science of no mean scope. It is a formidable body of knowledge much of which has no application whatever to accounting. This volume treats of those basic procedures applicable to business and accounting data generally. Many institutions such as banks and insurance companies require much greater elaboration and detailed analyses of their figures than do ordinary trading concerns, but a consideration of the advanced methods of statistics necessary for that, does not lie within the scope of this book.

Reverting to a consideration of the objects of statistics as applied to accounting, we can say that the purpose is actually twofold and consists of—

- (a) the obtaining of a statistical record (which is a record of many events of the same type, e.g., the sales per week of certain types of shoes);
- (b) the analysis of such record so that conclusions may be drawn from it.

Supplementary to this is the treatment of figures involving graphs and diagrams, which will be considered in detail later (see Chapter 6).

Value of Statistics

An exhaustive survey of the benefits, advantages and values to be derived from the application of statistical methods to accounting procedures will not be attempted here. It will serve our purpose at this stage to indicate the main values in broad terms and leave specific details to be illustrated throughout the text. In general, the chief advantages attaching to the use of statistics are as follow:

- (a) It requires the observance of rules as to classification of the data and its treatment generally, and thus is in accord with the accounting doctrine of consistency.
- (b) It provides for the analysis of past figures with the aim of investigating the causes of the resultant effects.
- (c) Supplementary to (b) it provides a guide for the future elimination of elements resulting in undesirable effects as exhibited by past figures, and likewise, renders possible the continuance of desirable elements.
- (d) It provides methods for ascertaining the trend of figures, or the direction in which a certain set of events is moving.
- (e) It may be used to ascertain whether a particular set of events is by nature continuous, seasonal, fluctuating, permanent, progressive or regressive.
- (f) It makes possible the comparison of two sets of events so

as to determine the degree of correlation between the two, i.e., whether changes in one are associated with changes in the other and if so, the degree of the relationship between the two sets of events.

Whereas the methods of statistics were once rarely if ever applied to accounting problems, by virtue of these values they have become today an important part of the functionary equipment of modern accounting technique.

Limitations of Statistics

Mention should be made, however, of certain inherent limitations attached to the science. If, for instance, a possible future estimate relating to certain events is arrived at by the application of statistical methods, such result will not take into consideration the effect of possible social and economic changes. These latter effects must be gauged by human judgment and the statistical result arrived at duly modified by such allowances. Such factors, however, do not detract from the merit of first obtaining as solid a foundation as possible upon which to base human judgment.

In some cases the statistical result may be impaired by the lack of information which may be impossible or too difficult to obtain. For example, the total time lost in a factory due to sick leave, may be known or arrived at, but it may not be possible to ascertain the causes of sickness so as to judge whether they were of a recurrent or non-recurrent type and thus would be likely or otherwise to account for a similar loss of time in the year ahead. In many cases the keeping of records to show such details would place too great a strain upon the administrative section of the business and the record must be shelved.

Another deficiency results from the fact that when dealing with large masses of items, small but important details may be swamped, and in the desire to obtain the general result, produce no effect. More will be said later as to the effect of extreme items in a series.

Internal Statistics

Internal statistics is a name given to the statistical analysis of data applicable to a particular business. It is the study of statistical data drawn from within the business or organization itself. The following give some idea of what kinds of internal statistics could be prepared by a business:

- (a) *In connection with Sales*
 - (i) Table of monthly average of sales for a year.
 - (ii) Table of comparative sales for a period of years.
 - (iii) Table of sales applicable to certain territories.
 - (iv) Table of comparative sales effected by salesmen.
 - (v) Table of comparative sales of the different departments of the business.

(b) In connection with Production

- (i) Table of cost prices of different commodities produced for a year or other period.
- (ii) Table of quantities of commodities produced for a year or other period.
- (iii) Table of comparative production figures for the different departments on basis of cost and/or quantity.
- (iv) Table detailing comparisons of component parts of cost of production (materials used, direct labour, prime cost, etc.) for certain periods.

The above are, of course, only an indication of the many different types of data which could be subjected to statistical analysis, and, further, it should be noted that various statistical devices may be applied to these tables. In some businesses, the collection and tabulation of data are required on a large scale, particularly when details relating to marketing, advertising, finance, etc., are also required. In those cases it is generally necessary to have a special statistical staff to work in conjunction with the accounting staff. As to whether such a special staff is warranted in any particular case will depend, of course, upon the size and nature of the business. Should it be necessary, then the respective functions of the accounting department and the statistical department should be well defined, and to what degree of control (if any) the former should exercise over the latter is a matter which the particular business must consider for itself. Where a separate statistical department is not warranted, it is the accountant who must perform the functions normally carried out by that department. (See also under Collection of the Data, *post.*)

The custom is growing yearly of presenting the accounts of a company to its shareholders in such fashion that the distribution of the turnover between materials, labour, taxation, reserves and dividends are shewn.

This is done either in total cash form as from the figures of I.C.I:

	1947 £m	1948 £m
Gross manufacturing and trading proceeds and gross income from investments, etc.	140 4	167 0
Raw materials and purchases for resale, maintenance of plants, freight charges, factory and sales administration expenses (exclusive of salaries and wages)	81 8	96·1
Wages and salaries	34·2	40·1
Pensions and contributions to pension funds	2·9	2·5
Obsolescence and depreciation of plants	4·3	5·3
United Kingdom and overseas taxation	8 9	11·1
Retained by the company for additions to reserves, including special addition to central obsolescence and depreciation provision	4 4	7·5
Distributed as net dividends to stockholders	3·9	4 4
	140 4	167 0

or as a percentage of the whole as in the Bristol Aeroplane Co. Ltd.'s accounts.

Here is the break-up of the company's income of 1948:

	Per cent.
Materials, power and services	48.25
Wages, salaries and employees benefits	42.87
Depreciation	2.28
Taxation, apart from P.A.Y.E.	2.56
Net profit	4 04
	<hr/> 100 00 <hr/>

or from the accounts of a famous Brewery:

	Per cent.
Beer and licence duties	69 0
Taxation	5.1
Materials and services	8.7
Wages and salaries	4 4
Trade expenses	6.6
Depreciation and reserves	4 2
Interest	.3
Dividend	1 7
	<hr/> 100 0 <hr/>

Another variant is to show the break-up of every £1 of turnover as in the figures for the Glacier Metal Company given below:

THE GLACIER METAL CO. LTD.
Financial Year 1947/48 (All Branches)
Allocation of Each Sales £1

Expenses	Parts to Nearest Farthing
	s. d.
PRODUCTION MATERIAL	5 4½
WAGES AND SALARIES	6 4½
GENERAL EXPENSES:	
(a) <i>Services and Supplies</i>	
Rent and Rates	¾
Gas and Electricity, Fuel Oil, Coke, Steam	4½
Consumable Supplies	6½
Stationery and Office Supplies	1
Carriage, Postage and Telephone	2
Insurance, National, Fire and General	2
(b) <i>Maintenance and Depreciation</i>	
Repairs and Renewals	1 2½
Depreciation	2½
(c) <i>Sundry Expenses</i>	
Advertising, Discounts Allowed, etc.	5½
(d) <i>Employees' Welfare</i>	
Training, Gratuities, Canteen, etc.	1½
Pensions	1
(e) <i>Scrap</i>	1 1½
(f) <i>Other Expenses</i>	
Patent Action, etc.	2½
PROFIT DISTRIBUTION:	
Taxes	1 9½
Shareholders' Dividends Less Tax	2
Reserves	1 6½
	<hr/> 20 0 <hr/>

External Statistics

External statistics consist of statistics released by the statistical departments of the various ministries and local government authorities, or by the United Nations and other international bodies, as well as by the Stock Exchanges, Banks and other institutions. They are often of great importance to a business in the matter of its organization, control and planning.

In addition, see later under Market Analysis for other aspects relating to External Statistics.

The following is a specimen of a table of statistics relating to the retail trade, extracted from the *Monthly Digest of Statistics* published by the Government.

Sales by independent traders⁽¹⁾

Weekly average 1947 = 100

	Drapers ⁽²⁾	Men's wear shops			Boot and shoe retailers	Furniture dealers	Hardware, electrical and radio shops		Chemists	
		Total	Out-fitters	Bespoke tailors ⁽³⁾			Total	Electrical and radio shops		
1947	100	100	100	100	100	100	100	100	100	
1948	117	119	120	118	128	111	101	103	101	
1949	130	130	134	125	128	124	100	111	106	
1950	129	135	137	133	129	138	102	122	107	
1949	April ...	142	143	147	139	138	134	102	119	104
	May ...	142	145	150	140	142	129	108	123	99
	*June ...	131	138	147	129	133	116	89	105	101
	July ...	127	144	145	144	134	122	98	99	115
	August ...	101	103	105	101	107	111	89	86	105
	*September	120	109	110	107	127	121	96	94	102
	October ...	153	141	144	137	155	151	112	132	100
	November	138	131	135	126	140	139	104	123	106
	*December	168	182	200	160	150	146	122	139	129
1950	January ...	113	110	110	110	99	118	96	128	103
	February ...	100	96	94	98	85	121	98	132	106
	*March ...	121	118	116	119	113	131	95	116	103
	April ...	123	126	129	122	121	147	96	117	111
	May ...	137	148	154	143	151	134	108	122	103
	*June ...	121	141	147	135	134	122	98	117	102
	July ...	125	135	139	132	128	130	97	102	115
	August ...	105	117	120	114	116	122	91	95	108
	*September	145	148	144	151	150	149	102	117	102
	October ...	146	146	140	152	149	162	109	137	97
	November	133	140	138	142	142	151	105	125	101
	*December	175	199	213	182	164	166	130	159	133
1951	January ...	121	135	130	139	106	138	106	140	111
	February ...	125	130	130	130	105	189	126	193	118
	*March ...	136	160	155	166	124	157	112	160	98
	April ...	138	145	148	143	137	178	133	230	106
	May ...	128	153	153	153	147	149	118	146	102
	*June ...	124	133	137	130	173	137	109	141	113

Source: Board of Trade.

(1) Details relating to the basis and size of sample, etc., were published in the *Board of Trade Journal* for November 1947.

(2) Including women's clothing shops of all kinds.

(3) Including all shops which do a substantial trade in made-to-measure garments.

* Average of five weeks.

It will be appreciated that the above table may be of considerable value to the management, if it prepares corresponding tables so that comparisons can be made between, say, the percentage increase or decrease of net sales of the business for the relevant periods with that as applicable to the trade generally, and so on.

External statistics are relied upon when organizing sales campaigns in that they supply population figures for districts or towns and thus quota systems can be devised for the distribution and sale of the firm's product. It may also be important to know whether particular districts are agricultural, pastoral, coal producing, mineral bearing, etc., districts, in that the product to be sold may have no market in any one or more of such districts. Here again the information is derived from external statistics.

These examples are an indication of the vast number of uses to which external statistics may be put.

There are a number of official publications available which supply valuable information of the nature indicated above, particularly, the Government Annual Abstract of Statistics, the Monthly Digest of Statistics, the Board of Trade and Ministry of Labour journals, the numerous white papers on national income and expenditure and other subjects published by the Stationery Office, the monthly reviews of the Banks, the Stock Exchange Official List, etc.

Part of a page from the latter is reprinted here to give an idea of the information it contains. It will be seen that this, together with other information contained in the Gazette relative to company meetings and details of accounts, etc., affords a valuable source of information generally, to accountants as well as to investors.

In addition, illustrations of external statistical diagrams are given in Chapter 6.

THE COLLECTION OF THE DATA

In general, the accountant or other person responsible for the statistical analysis in a particular organization, will be faced with a problem which will come under one of three categories:

- (a) It may be concerned with the statistical analysis of data to be derived *wholly from within the organization itself*, e.g., analysis of the departmental sales of the organization. (See under Internal Statistics, *ante*.)
- (b) It may be concerned with the statistical analysis of data to be derived *wholly from without the organization*. (See under Market Analysis, *post*.)
- (c) It may be concerned with the statistical analysis of data derived partly from within and partly from without the organization (e.g., analysis and comparison of the profits of the business over a period of years with those of the trade generally over the same period).

INVESTMENT TRUSTS

Marks received 126

When Last m	Dividends/Interest				NAME	Units of Ownership	QUOTATIONS Jan. 3	BUSINESS DATA
	Block	Current	Total for year	Total for previous year				
19 Nov.	3	2	..	7	A bbots Investment Trust, Lim, Ord. Stock	1	30 1/6 - 32 1/6	32 1/6 (12/25)
17 Oct.	4	2	..	11	Borden & Canadian Invest. Trust, Lim, Ord. Stock	1	50 1/2 - 56 1/2	52 1/2 (12/25)
17 Oct.	4	2	..	4	Do. do. 4% Cum. Pref. Stock	1	14 1/4 - 16 1/4	15 1/4 (12/25)
17 Oct.	11	11	..	35	Do. do. 3 1/2% Red. Deb. Stock (1928-33)	100	98 - 99	98 1/2 (11/18)
7 Aug.	41	34	..	41	Aberdeen, Clyn & Eastern Trust, Lim	1	13 1/2 - 15 1/6	13 1/6 (12/25) 12 1/6
17 Jan. 21	6	6	..	6	Aberdeen, Edinburgh & London Trust, Lim, Ord. (U)	1	22 1/2 - 27 1/6	25 1/6 (12/25)
28 Nov.	2	2	..	4	Do. do. 4% Cum. Pref. (U)	100	14 1/4 - 16 1/4	15 1/4 (12/25)
12 Dec.	11	11	..	34	Do. do. 3 1/2% Deb. Stock 1926-70	100	83 - 88	84 1/2 (12/25)
28 Nov.	2	2	..	6	Aberdeen Trust Co. Lim, 5% Red. Deb. Stock	100	97 - 102	101 (12/25)
7 Sept.	6	6	..	6	Alta Investment Trust, Lim, Ord. Stock	100	110 - 120	110 (12/18)
14 Nov.	2	2	..	5	Do. do. 5% Cum. Pref. Stock	100	94 - 99	97 (12/25)
12 Dec.	11	11	..	41	Alliance Investment Co. Lim, Cum. 4 1/2% Pref. Stock	100	85 - 90	86 (12/25)
3 Oct.	2	2	..	7	Do. do. Deb. Stock	100	105 - 107	106 (12/25) 16 1/2
11 Oct.	11	11	..	3	Do. do. 3% Deb. Stock, 1922-42	100	73 - 78	81 1/2 (12/25)
19 Sept.	15	15	..	40	Alliance Trust Co., Lim, Ord. Stock	1	94 - 104	103 1/2 (12/25)
19 Sept.	2	2	..	4	Do. do. 4% Pref. Stock	100	80 - 85	83 1/2 (12/25)
19 Sept.	2	2	..	4	Do. do. 4 1/2% Cum. Pref. Stock	100	83 - 88	84 1/2 (12/25)
19 Sept.	2	2	..	6	Do. do. 5% Pref. Stock	100	97 - 102	99 (12/25)
31 Oct.	2	2	..	4	Do. do. 4 1/2% Deb. Stock, Red. 1926	100	95 - 100	100 1/2 (12/25)
31 Oct.	2	2	..	4	Do. do. 4 1/2% Deb. Stk. Red. after 1925/26	100	94 - 99	96 (12/25)
22 Aug.	2	2	..	8	American Inv. & Gen. Trst. Co. Lim, 5% Nn-Cm Pfd. Stk	93 - 98	96 (12/25)	
22 Aug.	3	3	..	8	Do. do. Deb. Stock	100	198 - 208	200 (12/26)
31 Oct.	2	2	..	4	Do. do. 4% Deb. Stock, Red. 1927	100	85 - 90	88 1/2 (12/26)
31 Oct.	2	2	..	4	Do. do. 4 1/2% 2d Deb. Stock, Red. 1927	100	90 - 95	92 (12/25)
18 July	4	4	..	11	American Trust Co. Lim, Ord. Stock	100	240 - 260	248 1/2 (11/15)
18 July	2	2	..	6	Do. do. 5% Cum. Pref. Stock	100	96 - 100	98 (12/15)
19 Sept.	11	11	..	35	Do. do. 3 1/2% Deb. Stock, 1926-75	100	78 - 83	80 1/2 (11/15)
17 Oct.	6	6	..	17	Anglo-American Debenture Corp. Lim, Ord. Stock	100	346 - 355	352 1/2
1 Aug.	2	2	..	4	Do. do. 4% Cum. Pref. Stock	100	83 - 88	85 1/2 (11/25)
12 Dec.	2	2	..	4	Do. do. 4% Deb. Stock, Red.	100	83 - 87	85 (12/25)
14 Nov.	3	3	..	8	Anglo-Celtic Trust, Lim, Ord. Stock	100	147 - 157	156 1/2 (12/25)
18 Aug.	2	2	..	6	Do. do. 5% Cum. Pref. Stock	100	95 - 100	96 1/2 (12/25)
17 Oct.	4	4	..	6	Do. do. 5% Red. Deb. Stock	100	98 - 103	99 1/2 (12/25)
14 Nov.	10	10	..	10	Anglo-Scottish Investment Trust, Lim, Ord. Stock	100	265 - 275	270 1/2 (12/25)
14 Nov.	2	2	..	6	Do. do. 5% Cum. Pref. Stock	100	95 - 100	98 (12/25)
31 Oct.	11	11	..	13	Do. do. 3 1/2% Deb. Stock, 1922-46	100	85 - 90	86 1/2 (12/25)
2 Jan.	11	11	..	6	Army & Navy Invest. Trst. Co. Lim, 5% Cum. Pfd. Stk	93 - 98	94 (12/25)	
2 Jan.	11	11	..	14	Do. do. Deb. Stock	100	280 - 290	285 (12/25)
12 Dec.	2	2	..	4	Do. do. 4% Perp. Red. Deb. Stock	100	265 - 275	270 1/2 (12/25)
11 May	7	7	..	29	Ashdown Investment Trust, Lim, Ord. Stock	100	165 - 175	168 1/2 (12/25)
11 May	11	11	..	35	Do. do. 3 1/2% Cum. Pref. Stock	100	70 - 75	81 1/2 (11/25)
11 May	7	7	..	6	Associated City Investment Trust, Lim, Stock	21	21 - 26	24 1/2 (12/25) 2 1/2
4 June	4	4	..	3	Atlas Electric & General Trust, Lim, Ord. Stock	1	189 - 203 1/2	194 (12/25) 2 1/2
17 Oct.	3	3	..	7	Do. do. 7% Pref. (Cum) Stock	1	26 1/4 - 28 1/4	27 1/2 (12/25)
17 Oct.	2	2	..	6	Do. do. 5% Deb. Stock, Red.	100	95 - 100	101 1/2 (12/25)
31 Oct.	2	2	..	4	B ankers' Invest. Trust, Lim, 4 1/2% Pref. Stk. (Cum.)	100	85 - 90	88
31 Oct.	3	3	..	9	Do. do. Deb. Stock	100	118 - 125	122 (12/25)
12 Dec.	2	2	..	4	Do. do. 4% Perp. Deb. Stock	100	78 - 83	81 (12/25)
19 Jan. 31	6	6	..	6	Bay Hill Trust, Lim, Ord	1	36 1/2 - 40 1/2	40 (12/15)
31 Dec.	14	14	..	35	Do. do. 5 1/2% Ser. "A" Red. Debts, Red.	100	92 - 97	95 (12/25)
19 Feb.	2	2	..	4	Beaver Trust, Lim, 6% Cum. Pref. Stock (U)	100	185 - 195	190 (12/25)
14 Nov.	2	2	..	4	Brewery & Com. Invest. Tr. Lim, Pfd. Ord. Stk. (4% Cum)	100	78 - 83	80 1/2 (12/25)
31 Oct.	4	4	..	12	(British & Foreign General Securities & Investment) Trust, Lim, Ord. Stock	100	210 - 220	214 1/2 (12/25)
3 Oct.	2	2	..	5	Do. do. 5% Cum. Pref. Stock	100	95 - 100	96 (12/25)
12 Dec.	2	2	..	4	Do. do. 4% 1st Deb. Stock, Red.	100	90 - 96	94 1/2 (12/25)
14 Nov.	49200	26	..	45	British Assets Trust, Lim, Ord. Stock	57	43 1/4 - 46 1/4	45 1/6 (12/25)
8 Sept.	637 1/2	Do. do. 4 1/2% Pref. Stock (Cum.)	100	89 - 93	90 (12/25)
8 Sept.	637 1/2	Do. do. "A" 5% Pref. Stock (Cum.)	100	95 - 100	99 1/2 (12/25)
8 Oct.	2	2	..	6	Do. do. 5% Deb. Stock, Red.	100	98 - 104	102 1/2 (12/25)
3 Oct.	2	2	..	6	Do. do. 5% 50-Yr. Deb. Stock (1977)	100	98 - 103	103 1/2 (12/25)
14 Nov.	2	2	..	16	British Canadian Trust, Lim, Ord. Stock	100	320 - 340	330 (12/25)
19 Oct.	3	3	..	7	British Combined Investors Trust, Lim, Ord. Stock	100	136 - 148	149 (11/15)
12 Dec.	2	2	..	6	Do. do. 5% Cum. Pref. Stock	100	83 - 88	86 1/2 (12/25)
14 Nov.	2	2	..	6	Do. do. 5% 1st Deb. Stock, 1926-78	100	100 - 105	102 (12/25)
17 Oct.	4	4	..	7	British Empire Land Mt & In. Co. Lim, Ord. Stk	1	11 - 11 1/2	11 1/2 (12/25)
17 Oct.	4	4	..	7	Do. do. New Ord. Stock	1	11 - 11 1/2	11 1/2 (12/25)
4 July	6	6	..	6	British Ind. & Gen. Invest. Trust, Lim, 5% Cum. Pfd. Stk	100	65 - 70	68 (12/25)
11 Dec.	6	6	..	24	Do. do. Deb. Stock	100	40 - 45	43 1/2 (12/25)
31 Oct.	2	2	..	24	British Investment Trust, Lim, 3 1/2% Pref. Stk. (Cum)	100	95 - 100	96 1/2
31 Oct.	8	8	..	24	Do. do. Deb. Stock	1	77 1/2 - 82 1/2	80 1/2 (12/25) 80 1/2
8 Sept.	14	11	..	3	Do. do. 3 1/2% Deb. Stock, 1925-77	100	75 - 80	76 1/2 (12/25)
5 June	6	6	..	6	British Isles & Gen. Invest. Trust, Lim, Ord. Stock	1	18 1/2 - 19 1/2	18 1/2 (12/25) 18 1/2
5 July	6	6	..	6	Do. do. 4% Cum. Pref. Stock	100	85 - 101	93 1/2 (12/25)
2 Jan.	3	3	..	5	British Steamship Invest. Trust, Lim, 5% Cum. Pfd. Stk	100	105 - 110	107 (12/25)
2 Jan.	2	2	..	10	Do. do. Deb. Stock	100	175 - 185	180 1/2
1 Jan.	2	2	..	4	Do. do. 4 1/2% Perp. Deb. Stock	100	89 - 94	91 1/2 (12/25)
2 Jan.	2	2	..	4	Do. do. 5% Deb. Stock, Red.	100	96 - 103	99 1/2 (12/25)
28 Nov.	2	2	..	6	Broadsides Investment Trust, Lim, 5% Deb. Stk., 1923-58	100	90 - 100	97 1/2 (12/25)
20 Jan.	3	3	..	8	Drumner Investment Trust, Lim, Ord. Stock	100	105 - 110	107 1/2 (12/25)
12 Dec.	2	2	Do. do. 5% Cum. Pref. Stock	100	93 - 98	107 1/2 (12/25)
14 Nov.	7	7	..	8	C .L.R.P. Investment Trust, Lim, Ord. Stock	100	119 - 124	120 (12/25)
14 Nov.	2	2	..	3	Do. do. 5% Cum. Pref. Stk	100	91 - 96	92 (12/25)
8 Aug.	11	11	..	3	Do. do. 3% Deb. Stock, 1922-77	100	75 - 80	78 (12/25)
8 Aug.	2	2	..	6	Do. do. 4% Deb. Stock, 1922-77	100	90 - 96	100 1/2 (12/25)
22 Aug.	2	2	..	8	C.R.C. Investment Trust, Lim, Ord. Stock	57	41 - 45	43 1/2 (12/25)
1 Aug. 31	6	6	..	8	Cable & Wireless (Holding) Trust, Lim, Ord. Stock	1	115 - 119	117 1/2 (12/25)
12 Dec.	15	3	Do. do. 3 1/2% Unsecured Loan Stock, Red.	100	96 - 98	97 1/2
12 Dec.	2	2	..	4	Do. do. 4% 10-Yr. Unsecured Loan Stock, Red.	100	94 - 96	95 1/2 (12/25) 95 1/2
14 Nov.	2	2	..	4	Cables Invest. Trust, Lim, 4 1/2% Cum. Red. Pref. Stock	1	17 1/2 - 19 1/2	18 1/2 (12/25)
14 Nov.	11	11	..	38	Do. do. 3% Red. Deb. Stock	100	120 - 130	126 (12/25)
3 Oct.	5 (20)	Calcedonia Investments, Lim, 5% Cum. "A" Pref. Stock	1	29 1/6 - 31 1/6	29 1/6 (12/25)
19 Oct.	5 (20)	Do. do. 5% Cum. "B" Pref. Stock	1	27 1/2 - 29 1/2	27 1/2 (12/25)
19 Oct.	2	2	..	4	Calcedonia Trust Co., Lim, 4 1/2% Cum. Pref. Stock	100	83 - 88	85 (12/25)
14 Dec.	11	11	..	38	Do. do. 3% Red. Deb. Stock	100	85 - 89	86 1/2 (12/25)
31 Oct.	3	3	..	15	Camperdown Trust Co. Lim, Ord. Stock	1	37 1/2 - 42 1/2	40 1/2 (12/25)
31 Oct.	2	2	..	4	Do. do. 4% Cum. Pref. Stock	100	78 - 83	79 1/2 (12/25)
28 Nov.	8	8	..	10	Canadian & Foreign Invest. Trust, Lim, Ord. (U)	1	61 1/2 - 67 1/2	64 (12/25)
19 Sept.	2	2	..	8	Do. do. 5% Cum. Pref. (U)	1	219 - 229	225 (12/25)
14 Nov.	4	4	..	12	Capital & National Trust, Lim, Ord. Stock	100	97 - 102	99 (12/25)
2 Jan.	2	2	..	5	Do. do. 5% Cum. Pref. Stock	100	94 - 99	100 1/2 (12/25)
4 July	2	2	..	5	Do. do. 5% Red. Deb. Stock, 1928	100	99 - 104	104 (12/25)
12 Dec.	2	2	..	8	Cardinal Invest. Trust, Lim, 5% Cum. Pref. Stock	100	93 - 98	96 1/2 (12/25)
4 June	3	3	..	7	Do. do. Deb. Stock	100	131 - 136	134 1/2 (12/25)

Ex share Bona (1/12/25)

FINANCIAL TRUSTS, LAND AND PROPERTY—Continued

When Last at	Dividends/Interest				NAME	Unit of Quotation	QUOTATIONS Jan. 3	BUSINESS DAYS
	Rate % actual	Corrected Rate %	Total for year %	Total for previous year %				
12 Dec.	2 1/2	2 1/2	4 1/2	4 1/2	Tampkin's Licensed Proprietary, Ltd. 4 1/2% 1st Mt. Deb. S. K. Rd. at Co. Ltd.	100	96 - 99	100 (12/10/51)
17 Oct.	2 1/2	2 1/2	2	2	do.	(U)	3/6 - 4/6	4 (10/12/51)
31 Oct.	2 1/2	2 1/2	7	7	Tea Share Trust, Ltd. Stock	107	6/6 - 8/7	7/22 (10/12/51)
30 Nov. '49	11	11	11	11	Tocks (Argentine) Land Co., Ltd.	1	23.9 - 26/3	24/7 (11/11/51)
3 Oct.	3	3	8	8	Telephone & General Trust, Ltd. Ord. Stock	1	23.9 - 35/7	33/9 (3/11/12/51)
2 Jan.	3 1/2	3 1/2	7	7	do. do. 7 1/2% Cum. Pref. Stock	1	23.9 - 26/3	23.9 (11/12/51) 23/62
8 Aug.	1 1/2	1 1/2	3 1/2	3 1/2	do. do. 3 1/2% 1st Deb. Stock, 1967-72	100	99 - 99	99/4 (11/11/51)
12 Dec.	6	6	5	5	Thames Estates & Investments, Ltd. Ord. Stock	1/7	1/9 - 1/3	1/10 (1/11/12/51)
5 Sept.	2 1/2	2 1/2	5 1/2	5 1/2	do. do. 5 1/2% Cum. Red. Pref.	1	13/9 - 16/3	13/9 (11/12/51)
5 Sept.	10	10	39	39	Thames House Estate, Ltd. 2 1/2% 1st Mt. Deb. S. K. Rd. 1963	100	49 - 49	49/90 (11/12/51)
3 Oct.	2 1/2	2 1/2	4 1/2	4 1/2	"The Adelphi" (London), Ltd. 4 1/2% Cum. Pref.	1	17/6 - 20/7	17/6 (6/25/47/51)
(20th)	10 cts.	10	20 cts.	20 cts.	Third Canadian Gen. Invest. Tr. Ltd. Sbs. of no par val.
....	(Toronto & London Investment Co., Ltd. Common)
2 Jan.	10	10	15	15	Town Investments, Ltd. Ord. Shares of no par value	(U)	57/6 - 62/3	62 (11/10/51)
3 Oct.	3 1/2	3 1/2	7	7	do. do. 7 1/2% Cum. Pref.	(U)	23/1 - 25/1	23/7 (11/12/51)
17 Oct.	8	8	10 1/2	10 1/2	Trafford Park Estates, Ltd. Ord.	(U)	40/7 - 42/6	42/6 (11/7/51)
23 May	5	5	5	5	Transval Land Co., Ltd. Ord. Stock	2/7	2/3 - 3/3	2/3 (11/12/51)
23 May	6	6	6	6	Trust & Agr. Co. of Australasia, Ltd. Ord. (S. with 1st Mt. Deb. S. K. Rd.)	..	18/9 - 21/3	20 (5/9/51)
23 May	6	6	6	6	do. do. Ord.	9	8 - 9	8 (11/12/51)
2 Jan.	2	2	4	4	do. do. 4% Cum. Pref. Stock	100	66 - 73 1/2	68 (11/12/51)
12 Dec.	2 1/2	2 1/2	Trust & Mort. Co. of Iowa, Ltd. 5% Cum. Pref.	1	17/6 - 20/7	18/7 (11/10/11/51)
18 July	2 1/2	2 1/2	do. do. 4 1/2% Deb. Stock, Red.	100	80 - 90	93/8 (7/11/51)
26 Nov.	2 1/2	2 1/2	United City Property Trust, Ltd. 5 1/2% Cum. Pref.	1	18/1 - 20/7	18/3 (6/22/11/51)
19 Dec.	30 cts.	30 cts.	United Corps. Ltd., Class "B" Shares of no par value	..	7/6 - 8/1	7/8 (11/12/51)
12 Dec.	7 1/2	7 1/2	United Dominions Trust, Ltd. Ord. Stock	1	57/6 - 62/6	59 (8/10/51) 9/6 W
12 Dec.	2 1/2	2 1/2	do. do. 4 1/2% Cum. Pref. Stock	1	17/1 - 19/1	18 (12/11/51)
17 Oct.	7	7	United Motor Finance Corp., Ltd., 7% Cum. P. Pref.	1	19/4/5 - 21/10/3	19.6 (11/12/51)
23 May	27 1/2 P	32 1/2 P	27 1/2 P	32 1/2 P	Van Diemen's Land Co., "A"	5/7	22/1 - 34/7	22/6
2 Jan.	2 1/2	2 1/2	Victory Real Estate Co., Ltd. Ord. Stock	1/7	1/9 - 1/15	1/7 (11/10/51)
19 Sept.	2 1/2	2 1/2	Waddon Investments, Ltd. 4 1/2% Cum. Pref. Stock	1	17/6 - 20/7	18/9 (11/12/51)
2 Jan.	2 1/2	2 1/2	do. do. 6% Deb. Stock, Red.	100	95 - 100 1/2	97 1/2 (11/11/51)
6 Sept.	10	10	10	10	West African Mahogany Co., Ltd.	(U)	2/6 - 3/1	2/84 (11/12/51)
12 Dec.	2	2	4	4	{ West London Property Corp., Ltd. 4% 1st Mt. Deb. Stock, Red.	100	94 - 99	94
18 July	1 1/2	1 1/2	3 1/2	3 1/2	Western Ground Rents, Ltd. 3 1/2% 1st Mt. Deb. S. K. Rd.	100	76 - 81	79/90 (12/11/51)
2 Jan.	3	3	6	6	White House (Regent's Park), Ltd. 6% Cum. Pref. (U)	1	19/1 - 20/1	18/6 (11/12/51)
31 Oct.	6	6	8	8	Whitehall Court, Ltd. Ord.	(U)	6/1 - 8/1	6/6 (11/12/51)
8 Aug.	2 1/2	2 1/2	5 1/2	5 1/2	do. do. 5 1/2% Cum. Pref.	(U)	7/1 - 8/9	7/13 (11/51)
12 Dec.	1 1/2	1 1/2	do. do. 3 1/2% Mort. Deb. Stock, Red.	100	80 - 85	84 (11/7/51)
6 Sept.	1 1/2	1 1/2	{ Wool Exchange & Gen. Invests., Ltd. 3 1/2% 1st Mt. Deb. Stock (Red.)	100	75 - 80	80 (11/10/51)

GAS

Marks received 3

8 Sept.	3	3	..	6	{ Alliance & Dublin Consumers' Gas Co. (Alliance) and Dublin Gas Cons. Ord. Stock	1	20/6 - 22/6	22 (4/12/51)
31 Oct.	6 1/2	9	6 1/2	9	Bombay Gas Co., Ltd. Stock	1	18/1 - 23/4	23/9 (11/9/50)
31 Oct.	2 1/2	2 1/2	4 1/2	4 1/2	{ Cape Town & Dist. Gas Light & Coke Co., Ltd. 4 1/2% Cum. Pref. Stock	1	14/1 - 16/1	14/7 (7/12/51)
12 Dec.	2 1/2	..	5	..	do. do. 6% Deb. Stock, 1971 (London Register)	100	94 - 99	95 (12/11/51)
..	do. do. 5% do. (Dominion Register)	100	94 - 99	..
3 Oct.	3 1/2	3 1/2	8	8	Colombo Gas & Water Co., Ltd. Ord.	(U)	26/6 - 31/6	27/6 (12/11/51)
3 Oct.	3 1/2	3 1/2	7	7	do. do. Pref. (7% Non-Cum.)	(U)	21/1 - 25/1	22/1 (11/15/49)
19 Sept.	3 1/2	3 1/2	6	6	Colonial Gas Association, Ltd. Ord.	1	16/1 - 18/1	16 (10/6/51)
19 Sept.	4	4	8	8	do. do. 8% Pref.	1	20/1 - 22/1	19/9 (16/10/51)
6 Oct.	22 1/2 P	22 1/2 P	..	9	Imperial Continental Gas Assn., Capital Stock	100	115 - 120	117 1/2 (11/9/51)
12 Dec.	8 1/2	8 1/2	..	10	Malta & Mediterranean Gas Co., Ltd. Ord. Stock	100	113 - 118	117 (8/18/51)
12 Dec.	3 1/2	3 1/2	..	7	do. do. 7 1/2% Non-Cum. 1st Pref. Stock	100	113 - 118	117 (11/12/51)
12 Dec.	3 1/2	3 1/2	..	7 1/2	do. do. 7 1/2% Non-Cum. 2nd Pref. Stock	100	113 - 118	115 (11/12/51)
2 May	8	8	8	8	Montevideo Gas & Dry Dock Co., Ltd. Ord. Stock	100	60 - 70	63 1/2 (12/11/51)
3 Oct.	3	3	..	6	do. do. 6% 1st Mort. Deb. Stock, Red.	100	58 - 100	99 (11/10/51)
24 Dec. 46	2 1/2	4	2 1/2	4	Primitiva Holdings, Ltd. Ord. Stock	1	6/1 - 8/1	7/13 (11/12/51)
12 Dec.	3	3	6	6	do. do. 6% 1st Cum. Pref. Stock	100	93 - 98	95 (11/12/51)

INSURANCE

Marks received 8

When Last at	Amount of Ratio % actual	Corrected Amount of Ratio %	Amount for previous year %	Total for previous years	NAME	Unit of Quotation	QUOTATIONS Jan. 3	BUSINESS DAYS
14 Nov.	10/1	10/1	..	20/1	Aliance Assur Co., Ltd. Orig. Sbs. (L&D, with 2 1/2% p.d.)	(U)	24 - 25	24
17 Oct.	10/1	10/1	..	20/1	do. New Shares	(U)	24 - 25	24 1/2
2 May	16/6	4/6	12/9	12/9	Atlas Assurance Co., Ltd. (145, with 1 1/2% p.d.)	(U)	14 1/2 - 15 1/2	15 1/2 (7/12/51)
2 May	16/6	1/6	1/9	1/6	Bacon Insurance Co., Ltd. (116 1/2 with 5/16 p.d.)	1	37 - 40	37 (11/12/51)
12 Dec.	5 1/2	4 1/2	..	11 1/2	Islamic Assurance Co., Ltd. Ord. Stock	1	21 - 22	22 1/2 (11/12/51)
17 Oct.	1/9	1/6	3/6	3/1	do. 5% Tax-free Cum. Pref. Stock	1	36 1/2 - 40	36 1/2 (11/11/51) 35 1/2
17 Oct.	1/9	1/6	3/6	3/1	Calcedonian Insurance Co. (121, with 10/16 p.d.)	1	77 1/2 - 85	80 (11/12/51)

It is apparent, from this summary, where the person making the analysis will collect the data he requires. That is, the sources of his data may be wholly within the business, wholly without, or partly within and partly without. Reference has been made as to the nature of internal statistics and in this and the next succeeding six chapters, illustrations will be given of how problems relating to internal statistics are handled.

The most difficult part of collecting data is that of collating information supplied wholly from outside the organization. This problem concerns Market Research and a short description of its purposes and methods will now be given.

MARKET RESEARCH

Whilst Market Research normally does not come within the scope of the accountant of a particular organization, it is nevertheless essential that he should have some knowledge of its purposes and methods, not only that he may be able to interpret the figures it produces but also that he may appreciate the value to his organization of such research.

Quite often, the cause of fluctuations in such matters as turnover costs, absenteeism and unit production, are to be found outside, and *not* within, the manufacturing or distributing organization itself, and only by discovering the root causes can sound decisions or corrective action be taken.

Market Research, in its widest sense, has been authoritatively defined as "The study of all problems relating to the transfer and sale of goods and services from producer to consumer."

Some Market Research is of a general character, and has to do with marketing in its broad economic or social sense.

The great majority of marketing studies are, however, specific to individual businesses, and practically every normal marketing problem can be related to one or more of the following categories, whether they arise before, or after production:—

Studies relating to Policies, such as:

- (a) What amount of capital should be provided for promoting the sale of a product? Should the existing allocation be increased or reduced?
- (b) Should the product be sold outright or leased? Should it be sold for prompt payment or on instalments? Should the existing policy be varied in its entirety or in certain limited respects only?
- (c) The composition of the product itself—should it be made of metal, wood, rubber composition, etc.? Should its present composition be changed to another material such as a suitable plastic?

- (d) The form in which the product is to be offered—in bulk or in packages? Should the existing policy be varied?

If policies are to be soundly formulated, they must be based on facts, determinable only through competent research.

Studies relating to a Product include factors such as:

- (a) Its quality.
- (b) Its shape, size and composition.
- (c) The style, type and size of package.
- (d) The price at which it is to be sold.

All of the many decisions which have to be made involve considerations relating not only to the product itself but also relating to other products in competition with it. Obviously it is only with the aid of competent research that decisions can be soundly based.

Studies relating to the Market include factors such as:

- (a) The type and character of potential purchasers and their ability and disposition to buy.
- (b) The number of potential purchasers in different areas, and the relative accessibility of areas.
- (c) Marketing facilities.
- (d) The sales, advertising and service activities of competitors.

Again, research is of vital importance in gathering all the necessary market data.

Studies relating to Methods and means of Transfer and Sales, such as:

- (a) Sales Promotion.
- (b) Advertising and Selling.
- (c) Channels of Distribution.
- (d) Services required by competitive conditions.

Here again, it is necessary to have reliable data which can only be made available as a result of research.

Competent Market Research involves the following procedures:

1. The Analysis of the Problem.
2. The Planning of the Procedure.
3. The Collection of the Data.
4. The Organization of the Data.
5. The Interpretation of the Data.
6. The Presentation of the Results.

These procedures are best illustrated by briefly describing an actual example of detail required by an organization and how it would be handled.

1. The Analysis of the Problem

A manufacturer was considering the advisability of extending his activities in respect of certain of his proprietary brands and,

upon analysis, it was found that the problem resolved itself into the necessity for ascertaining, in relation to his own brands and all competitive proprietary brands and in a form which would enable comparisons to be made between all proprietary brands, the following detail:

- (a) The relative effectiveness of all forms of publicity and influences in inducing people to try each brand.
- (b) The relative effectiveness of each brand in maintaining continuous use in homes.
- (c) The brands now used most regularly by people, analysed into men, women, young children and other children.
- (d) The foregoing requirements analysed into types of brands.
- (e) Differences which exist, if any, as between metropolitan and country people.

2. The Planning of the Procedure

The first step was to find out how many different proprietary brands were currently on the market. This was ascertained, and totalled twenty-five different brands.

The second step was to design a suitable questionnaire on which could be recorded the desired information. The important matters to be considered in the preparation of the form were, (a) the information had to be recorded in the shortest possible time for interview, and (b) the information had to be capable of ready analysis.

The third step was to determine how the information was to be gathered, i.e., whether by

- (a) Telephone enquiries.
- (b) Mailed questionnaire.
- (c) Personal interview with people, i.e., in the street, in their offices, in other public places, etc.
- (d) Personal interviews with people in their homes.

Method (d) was adopted for the reasons, amongst others, that telephone homes did not represent a true cross section of all people; the questionnaire would appear too complicated to people who might receive it through the post; and the subject of the investigation was of too personal a nature to interrogate people in the streets, etc.

The fourth step was to pre-test the questionnaire. This was done by trained interviewers who conducted "Pilot" interviews and analysed the information they obtained.

/

3. The Collection of the Data

Trained interviewers were then briefed, and visited people in their own homes. The interviews were as far as possible properly spread both as to geographic area and social strata.

4. The Organization of the Data

Two methods of organizing the data were available:

- (a) The machine or mechanical method.
- (b) The manual or hand method.

Owing to the multiplicity of the individual items, it had previously been determined to use the hand method.

Extraction and analysis summary sheets had already been designed and prepared during the "Planning" stage, and each item of information for each completed questionnaire, after editing, was extracted and transferred to the appropriate extraction sheet.

Finally, all extraction sheets were summarized and converted to appropriate percentages, and tabulated.

5. The Interpretation of the Data

The tabulations were then considered separately in detail by three senior research executives, each of whom prepared a statement setting out his interpretation of the tabulations.

The three executives then met in conference, and the interpretation of the tabulations was finalized.

By courtesy of Sales Research Services Ltd. of London, there is reproduced below a series of tables illustrating part of a research into the market for quick frozen foods. Class AB are middle-class households whilst Class C are working-class.

Of a total of 5,002 housewives interviewed 630 or 12.6 per cent. stated they had purchased quick-frozen foods. This fairly low degree of purchase indicates what a large market yet remains to be cultivated.

Consumers' attitude to Quick-Frozen Foods

The 630 housewives who had purchased quick-freeze foods were asked why they bought these goods, and their answers are tabulated in Table I below:

TABLE I
Why Housewives Buy Quick-Frozen Foods

	Total	Great Britain	
		AB Class	C Class
Total Buyers	100 0	100 0	100 0
No particular reason, but like them ...	36.1	36.1	36 1
Very good for change and for special occasions.	19 6	21.2	18 4
Easy to prepare, very convenient and clean to use.	15.1	17 1	13 5
Quick, saves time, handy for emergency	9 7	11 2	8 5
Something new, have only tried once	9.2	4.9	12 3
Good in winter and when fresh vegetables are scarce.	3 9	3 4	4 1
Economical, no waste good value ...	2.1	2.1	2 1
Nice flavours, very like fresh ...	1.9	2.7	1 2
Better than dried or tinned goods ...	1.3	.6	1 8
Stock in own shop5	.7	.3
Can't get tinned3	—	.6
Nice for children3	—	.9

It is clear that consumers have not yet developed a real buying consciousness about these goods; to the mass, they are as yet novel, something new, attractive and nice. As will be shown later, a consistent frequency of purchasing has not yet fully developed, so that the important motives of purchasing behaviour, i.e. economy, no waste, convenience, etc. have not yet developed to their fullest extent into the field of buying idea-associations.

Over one-third of the housewives could not form a real reply to the question of why they bought, beyond the fact that they had tried frozen foods and liked them. The second largest proportion of reasons, nearly one-fifth (19·6) is still linked with novelty—good for a change and for special occasions. The inherent convenience of these goods, so appealing in these days of harassed housewives, only emerges in a quarter of the cases noted, where 15 per cent. (15·1) and 10 per cent. (9·7) referred to their ease of preparation, their convenience and cleanness in use, their time-saving qualities and so on. Moreover, it is interesting to note that extremely few of these users referred to the freshness and the preservation of the essential flavour in quick-frozen foods—only 2 per cent. (1·9).

Copy appeal based on the selling of vegetables and fruit out of season is not necessarily the best angle of approach and it would seem that content of propaganda would be better directed to educational work pointing out the economies of buying quick-freeze foods (as contrasted with the wastes occurring in the preparation of other foods) and the important quality of freshness which is an essential feature of these products.

Frequency of Purchase

“About how often do you buy quick-frozen foods?” was the next question put to the 630 housewife buyers, and their replies are shown in Table 2.

TABLE 2
Frequency of Purchase of Quick-Frozen Foods

	Total	Great Britain AB Class	C Class
Total Buyers	100 0	100 0	100 0
Have bought occasionally only ...	23 8	25 3	22·5
Have bought once or twice only ...	18 5	13 5	22·9
About once in three months or more	5 4	4·1	6·5
About every one or two months ...	17 4	13·8	20·1
Every two or three weeks	13·7	13·9	13·4
Once or more times per week ...	21·2	29·4	14 6

Thus, about 40 per cent. of these housewives must be regarded as “rare” users, i.e. the occasionals.

Taking all those who do not buy more frequently than every two or three weeks, it can be said that nearly two-thirds are just chance or occasional buyers, whilst just over one-third (34·9 per cent.) are

developing a genuine buying habit. But the 21·2 per cent. who do buy once or more times per week, represent only 3 per cent. (2·7) of the total sample. So there is quite a way to go.

What kinds of Frozen Foods are bought

The general availability of these products in this country, is largely vegetables and fruit, but increasingly more goods are coming into the frozen food supply—meat, fish, poultry, made-up dishes, etc.

Table 3 shows the products bought, and it will be seen that there is a fair degree of duplication in the purchasing.

TABLE 3
Types of Frozen Food Bought

Total Bought	Buying	100 0
	Vegetables	94 3
	Fruit	46 7
	Fish	2 2
	Meat	1 1
	Miscellaneous	1 0

The low proportions of the latter three items merely reflects, of course, the lesser availability of these types.

Why Housewives do not buy Frozen Foods

Asked "Why do you not buy Quick-Freeze Foods?" the following replies were given:

TABLE 4
Why Housewives do not buy Frozen Foods

Reason	Total	Great Britain AB Class	C Class
Total not buying Frozen Foods	100 0	100 0	100 0
They are too dear, not good value	46·6	30 2	49 4
No particular reason	14·6	20·5	13 6
Never had, never heard of them	10·2	8 3	10 6
Prefer fresh foods, don't keep	9·3	16·9	7 9
Don't want them, no use for them, plenty of other foods.	8·9	9 9	8 6
Don't like them, don't look appetizing, tried once, could taste carton.	7 0	8 4	6 8
Can't get easily	1·6	2 0	1·6
Does own preserving	·6	1·2	·5
Packets too large	·3	·7	·2
Has no refrigerator	·3	1 2	·2
Has refrigerator	·1	·3	·1
Grow own vegetables	·1	·3	·1
Doctor forbids	·1	·1	·1
No good for children	·1	—	·1
Too cold	·1	—	·1
Too messy	·1	—	·1

Resistances based on price figure largely in this list of reasons, covering one half of the working-class housewives and nearly one half of the total. One-sixth of the housewives could not form a reason, and a tenth had never heard of or tried them.

Of the remaining reasons showing any degree of intensity, most of them betray a lack of understanding of the selling points of these foods. For example, 9 per cent. said they did not buy frozen foods because they preferred "fresh foods" thus betraying ignorance of the quick-freezing process and of how it does in fact preserve the freshness of foods so treated. The next largest reason has something of the same intention—there are plenty of other foods, why should I buy these?

The fact that 7 per cent. said that they didn't like these foods, that they didn't look appetizing etc., may indicate the fact of badly kept installations or poor display methods. Although of only small dimensions the fact that 0.3 per cent. gave as a reason for non-purchase the fact that they had no refrigerator again shows a lack of understanding of the advantage of the product.

Broadly, this evidence links up with that shown earlier—there is a considerable degree of non-acceptance due to ignorance indicating the need for educational propaganda.

6. The Presentation of the Results

The draft report was prepared by the senior research executive in personal control of the investigation, and was then finalized at a further conference of the three executives.

The above outline briefly indicates the procedure adopted by Market Research in the determination of a particular problem.

THE ARRANGEMENT OF THE DATA

The Statistical Table

Having collected the data required, it becomes necessary to arrange or classify it so that it may be used for the particular purpose or purposes for which it was collected. There are many ways of tabulating the data collected and the nature and extent of the information required from the material will largely govern the method of presenting it. Generally speaking, a statistical table should conform to the following requirements:

- (a) It should be clearly titled so that the nature of the information it supplies is obvious.

- (b) It should indicate the source or sources from which the information was compiled.
- (c) It should be as concise as possible and not unwieldy. In some cases it may be better to use a summary table and attach the supporting tables.
- (d) Where particular items are subject to an exception or are otherwise modified, reference should be made to these by means of footnotes to the table, e.g. in a table which lists various companies with their respective figures for net profit say for the year ended 30th June, 1950, the financial year of one of the companies may have ended on 31st March, 1950. In that event an asterisk or other mark should be noted against that item and explanation given by way of a footnote.
- (e) The whole table should be arranged so that it supplies the information desired to be derived from the data.
- (f) In throwing into relief the different parts and features of the table, judicious use should be made of sub-headings, spacing of columns, different types, and underlining.
- (g) If possible columns and rows should be numbered so that reference to the table may be facilitated.

Parts of the Table

In general, a statistical table consists of the following parts:

- (a) The Title. Invariably it will be found that it is necessary to break this up into two parts, consisting of the main title and a sub-title.
- (b) An indication as to the source from which the data was compiled.
- (c) Captions, or headings for the various columns.
- (d) Designations for each of the rows. Should the rows be numerous it is desirable to number them. Such designations of rows are usually termed "stubs".

The above are really the basic parts of a statistical table and it may be necessary to include others, depending upon the particular table being prepared. For example, it may be necessary to add totals and sub-totals; the table may be lengthened so as to include averages, deviations, etc.; where the vertical columns include units which represent certain amounts then the heading to the table should state the designation of such units, and so on.

The following table sets out the various parts referred to above.

The Alpha Co. Ltd.

Title → Components of Factory Cost
 Sub-Title → For 5 years ended 1951.*
 Source → Source: Factory Cost Records
 Designation of Units → In £'000's

Column Headings or Captions →	Year	Materials	Labour	Expense	Total
	1947	28	9	4	41
	1948	30	10	5	45
	1949	27	10	4	41
	1950	32	11	6	49
	1951	35	12	7	54
	Total	152	52	26	230

Stubs → (points to the years 1947-1951)

Column → (points to the columns Materials, Labour, Expense, Total)

Horizontal Rows → (points to the rows for years 1947-1951)

Footnote → *Each year corresponds to the accounting year, viz : 1st July to 30th June.

STATISTICAL SERIES

A statistical series (or statistical distribution) is a systematic arrangement of numerical data prepared so as to effect analysis of the data contained therein. It is, in other words, the arrangement of the main data within the statistical table itself.

Two features are exhibited by the statistical series, viz. :

- (a) the *characteristic*, and
- (b) the *frequency*.

The *characteristic* is really the item which is capable of being measured or whose presence or absence can be noted, and the *frequency* is the measurement, value or quantity attaching or relating to such item. For example, in the frequency distribution illustrated below, the characteristic is the size of each shoe and the frequency is the quantity of shoes manufactured.

The different types of statistical series may be classified as follows :

Frequency Distribution

In this type of series the data is grouped according to magnitude or size (i.e. the *characteristic* is ranged in order of magnitude or size). A simple form of such a series would be as follows :

Size of Shoe	Quantity Manufactured
3	200
3½	220
4	70
4½	85
5	100
5½	400
6	650

This type of series is dealt with in detail later.

Time Series

Here, the data is grouped according to time of occurrence. The characteristic is measured regularly at intervals of a year, month, week or day. The following is a simple illustration of a time series:

Month	Sales £
January	900
February	756
March	843
April	926
May	543
June	432
July	376
August	248
September	598
October	641
November	859
December	946

The measurement of *Trend* (the direction in which the data is moving) in a Time Series is of great importance and will be dealt with in detail later. (See Chapter 5.)

Spatial Distribution

In this type of series the data is grouped according to geographical location. This type of classification automatically arises with certain matters, e.g. sales of a product in various districts. Such a series is normally arranged as follows:

District	Sales £
Birmingham	846
Newcastle	732
Manchester	700
Liverpool	628
Leeds	571
Plymouth	233

Category Series

Here, the data does not lend itself to classification according to size (frequency distribution), time (time series) or location (spatial distribution). It may simply consist of a list of items, as in the

following table, setting out a list of manufacturing companies with their respective net profits, for a particular year.

Company	Net Profit £
A.B.C Cotton Co. Ltd.	42,000
Manchester Export Co. Ltd.	31,500
K. Young & Co. (Oldham) Ltd.	10,250
Preston & District Weavers Ltd.	17,125
J. Flynn & Co. Ltd.	9,370
F. J. Mears Spinning Co. Ltd.	4,231

In the above classifications of statistical series we have only been concerned to point out the fundamental basis upon which each is prepared. It should be remembered that a particular series would be contained in its statistical table with appropriate title, sub-title, source, designation of units and so on as described earlier. In addition, it is as well to note here that the table as ultimately presented would generally carry the details of the analysis to which the data has been subjected. It may indicate the average of the series, standard or other types of deviation, degree of correlation, etc. These matters are considered later.

FREQUENCY DISTRIBUTION

The accountant constantly meets with this type of series and it is essential that he be familiar with the various forms it may take.

It will, perhaps, be best if we take an actual illustration of how data in its unorganized state can be transformed into an orderly series capable of being analysed.

Assume that the cost clerk of a manufacturing company has extracted from the alphabetical list of wage cards of employees of the factory, the following list of the different wages paid for a particular week.

£ s. d.	£ s. d.	£ s. d.
7 2 6	4 1 3	8 1 7
4 9 3	5 16 9	3 17 6
6 7 5	7 9 4	6 6 3
5 12 6	4 0 6	5 11 8
9 14 7	8 3 3	7 9 9
8 4 8	7 15 -	6 10 6
6 14 9	6 11 3	5 19 6
9 19 6	5 19 6	6 2 9
8 4 3	6 1 6	6 10 -
5 11 6	6 12 -	6 6 -
4 1 9	6 17 6	5 19 6
7 8 -	7 - -	6 15 9
5 7 6	6 9 9	6 11 6
5 17 6	6 - -	5 18 6
7 3 6	8 1 -	6 9 6
4 12 9	3 17 6	7 2 -
1 19 6	6 3 -	2 7 -
6 17 6	7 2 6	5 1 9
7 18 -	4 15 6	3 3 -
6 11 6	3 2 9	5 18 6
8 19 9	7 14 6	8 10 6
4 4 6	5 19 -	4 11 9
7 10 -	9 9 6	4 7 6

£ s. d.	£ s. d.	£ s. d.
5 9 9	6 3 6	7 2 9
6 7 3	4 1 7	8 9 -
1 17 6	3 15 9	7 8 6
7 2 6	5 16 9	3 17 6
8 3 9	9 19 6	4 - 6
7 2 6	5 7 7	8 9 -

There are 87 employees in the factory and all are represented by their respective wage in the above list.

It is obvious that the data as it is, is not suitably arranged for statistical purposes, and is not of much practical use while it exists in such an unorganized state. The first step to be taken to organize it is to decide upon suitable class intervals into which the different wages can be divided. A *class interval* covers a conveniently sized group forming part of the data as a whole. Numerically, it constitutes the difference between the highest value in the class below the group in question and the lowest value of the class above. We will assume that in this particular case, class intervals of £0.99 have been decided upon. Actually, in practice it will be found that experience will guide as to what is the best class interval for particular data for a given concern. Next, the highest and lowest figures are ascertained and the class intervals arranged in a column starting with the lowest first. Care should be taken to see that the class intervals do not overlap. A "score sheet" for the purpose of marshalling the different wages into their respective class intervals may now be prepared as follows:

X.Y.Z. COMPANY LIMITED

Score Sheet

Distribution of Employees
on basis of
Weekly Wages

Source: Employees' Wage Cards—27th July, 19.....

Weekly Wage	Tally	Frequency
*£1—£1.99	11	2
2— 2.99	1	1
3— 3.99	1 	6
4— 4.99	 	11
5— 5.99	 	16
6— 6.99	 	21
7— 7.99	 	16
8— 8.99	 	10
9— 9.99	1111	4
		—
		87
		—

* The respective class intervals are to be taken as reading—from and including £1 to and including £1.99, etc. The class interval is thus £0.99, and, it will be noted the class intervals as listed do not overlap.

It should be observed that it is not always necessary to arrange the characteristic into class intervals. In many cases it is of the simple form as illustrated on page 19, where the frequency is applied to a characteristic represented by a *particular size*, as distinct from a characteristic represented by a *particular class interval*, shown above.

In the tally the diagonal line crossed through each group of 4 units is counted as one and the total is thus 5.

Cumulative Frequency Distribution

The series, instead of being set out as showing the simple frequencies, may show the cumulative frequencies. The distribution of employees on the basis of weekly wages as shown in the previous table may have an added column showing the cumulative frequency distribution as follows:

X.Y.Z. COMPANY LIMITED**Distribution of Employees
on basis of
Weekly Wages**

Source: Employees' Wage Cards—27th July, 19...

Weekly Wage	Frequency	Cumulative Frequency
£1 - £1 99	2	2
2 - 2 99	1	3
3 - 3 99	6	9
4 - 4 99	11	20
5 - 5 99	16	36
6 - 6 99	21	57
7 - 7 99	16	73
8 - 8 99	10	83
9 - 9 99	4	87
---	---	---
	87	

Another form which the table may take is as follows:

X.Y.Z. COMPANY LIMITED**Distribution of Employees
on basis of
Weekly Wages**

Source: Employees' Wage Cards—27th July, 19...

Weekly Wage	Cumulative Frequency
£10 and over	0
9	4
8	14
7	30
6	51
5	67
4	78
3	84
2	85
1	87

As pointed out there are numerous ways in which the accumulated frequency may be arranged. For example the series could have been put into reverse as follows:

Weekly Wage	Cumulative Frequency
Under £2	2
.. £3	3
.. £4	9
.. £5	20
.. £6	36
.. £7	57
.. £8	73
.. £9	83
.. £10	87

Percentage Frequency Distribution

An extremely useful device is that of inserting an additional column into the ordinary table with its frequency distribution and show therein the percentage by which each item is represented in the total quantity. The table with this added feature would appear as follows:

X.Y.Z. COMPANY LIMITED

**Distribution of Employees
on basis of
Weekly Wages**

Source: **Employees' Wage Cards—27th July, 19...**

Weekly Wage	Frequency	Percentage
£1—£1.99	2	2.30
2— 2.99	1	1.15
3— 3.99	6	6.90
4— 4.99	11	12.64
5— 5.99	16	18.39
6— 6.99	21	24.13
7— 7.99	16	18.39
8— 8.99	10	11.50
9— 9.99	4	4.60
	87	100.00

The percentages supplied by this column are particularly important where two frequency distributions are ranged side by side and it is desired to compare them.

Another device is to have a percentage column showing the *percentage increase or decrease* of each item as compared, say, with the corresponding period last year.

CHAPTER 2

THE AVERAGE

INTRODUCTION	24
	Types of Averages	
THE ARITHMETIC AVERAGE	25
	Simple Arithmetic Average.	
	Illustrative Example 1	
	Short Method of Calculating Arithmetic Average.	
	Illustrative Example 2.	
	Weighted Arithmetic Average	
	Illustrative Example 3.	
	Averaging where Class Intervals used	
	Arithmetic Average and Trend	
	Advantages of the Arithmetic Average	
	Disadvantages of the Arithmetic Average.	
THE MEDIAN	31
	Illustrative Example 4.	
	Illustrative Example 5.	
	Illustrative Example 6.	
	Comparisons of Methods of Finding Median.	
	Advantages of Median.	
	Disadvantages of Median.	
THE MODE	35
	Illustrative Example 7.	
	Series with Class Intervals.	
	Illustrative Example 8	
	Advantages of the Mode	
	Disadvantages of the Mode.	
THE GEOMETRIC MEAN	38
	Illustrative Example 9.	
	Geometric Mean and Price Indexes.	
	Illustrative Example 10.	
	Advantages of Geometric Mean.	
	Disadvantages of Geometric Mean.	
THE HARMONIC MEAN	40
	Illustrative Example 11.	
	Advantages of the Harmonic Mean.	
	Disadvantages of the Harmonic Mean.	

INTRODUCTION

The *Average* is one of the basic devices used in the analysis of data. It may be defined as being the figure or item which tends to describe, or sum up, or represent, the distribution or series being considered.

In many cases the statistical analysis of data begins by finding one of the different types of averages discussed below and using that as the basis upon which to carry out further work.

Generally speaking, we can say that the following features are associated with the average:

- (a) The average is representative of the whole series.
- (b) Different series may be compared to a certain extent by comparing their averages.
- (c) The average is used as the basis of further analysis of the series.
- (d) It is a measure of central tendency.
- (e) Much labour is saved by substituting one figure or item for a number of figures or items.

From these points it will be realized why it is that the average plays such an important part in statistics. Actually, of the features noted above, not all are applicable to each and every one of the different types of averages described below, but modification if necessary will be made when considering the particular average itself.

It is also necessary before proceeding further, to point out that averages must be handled properly otherwise they may prove extremely dangerous by giving a false impression of the series. In addition it must be noted that an average does not reveal *trend*, i.e. an average does not show the *direction* in which a particular set of events may be moving.

Types of Average

There are five main types of average generally in use, viz.:

- (a) The *Arithmetic Average* } Symbol M_a .
(or *Arithmetic Mean*) }
- (b) The *Median*. Symbol M_e .
- (c) The *Mode*. Symbol M_o .
- (d) The *Geometric Mean*. Symbol M_g .
- (e) The *Harmonic Mean*. Symbol M_h .

The characteristics, advantages and disadvantages of each of these types will now be considered in detail.

THE ARITHMETIC AVERAGE.

Simple Arithmetic Average

The simple Arithmetic Average is calculated by adding together all of the items in the series and dividing such total by the number of items.

Illustrative Example 1

Assume that it is desired to find the Arithmetic Average of the Sales of the different types of cigarettes as per the following table.

THE DRAW-BACK TOBACCO CO. LTD.

Sales of Cigarettes by Types
Month Ended 30th April, 19...

Type No.	Sales £
1	600
2	350
3	900
4	1,200
5	700
6	800
7	850
8	1,000
9	800
	<hr/> 7,200 <hr/>

Here then, the Arithmetic Average is calculated by adding all of the sales values together to give £7,200 and dividing such total by the number of items (9), to give £800.

To express this calculation by way of a formula—

- Let M_a Arithmetic Average.
 N The number of items.
 x The data expressed as individual items (this is sometimes termed the "variate.")

The general formula is—

$$M_a = \frac{\sum(x)}{N} \text{ (the symbol } \sum \text{ called Sigma denotes "the sum of").}$$

In the above illustration the evaluation of the formula could be expressed as follows—

$$M_a = \frac{600 + 350 + 900 + 1200 + 700 + 800 + 850 + 1000 + 800}{9} = \frac{7,200}{9} = 800.$$

That is, the formula—

$$M_a = \frac{\sum(x)}{N}$$

means that the Arithmetic Average is equal to the sum of the individual items of the series divided by the number of the items.

It will be apparent from an examination of the above calculation that there is a feature attached to the Arithmetic Average which may be quite useful in practice. It is that if you know the Arithmetic Average of a series of items and know the number of items in the series, then you can ascertain the *total* value of the items. Again, if you know the total value of the items and also the Arithmetic Average, then you can ascertain the *number* of items in the series. In other words, of the three items, viz.: Arithmetic Average,

number of items, and total of items, given any two of them you can find the third.

Short Method of Calculating Arithmetic Average

There is a short method of calculating the Arithmetic Average which may be quite useful where the numbers are large. The procedure is to select an assumed average (known technically as the "Gussed Mean" or the "Arbitrary Origin" or "Assumed Mean") and obtain the deviation of each of the items from such Gussed Mean. These deviations are then added and divided by the number of the items to obtain the average deviation from the Gussed Mean. Such average deviation when ascertained is added to or subtracted from the Gussed Mean (according to whether the average deviation is positive or negative) and the resultant is the true Arithmetic Average.

The short method of calculating the Arithmetic Average may be illustrated by a simple example as follows:

Illustrative Example 2

The following series shows the number of motor-car tyres sold by a certain company for the relevant years. Calculate the Arithmetic Average by the Short Method of working.

Year	Number of Tyres Sold
1	1,250
2	1,100
3	1,300
4	1,350
5	1,500

Let the Gussed Mean be say 1,200.

Then the respective deviations of the tyres sold from the Gussed Mean will be as follows—

Year	Deviation
1	+ 50
2	- 100
3	+ 100
4	+ 150
5	+ 300

the sum of such deviations being + 500, which represents an average deviation of + 100 per year.

Therefore the true Arithmetic Average is the Gussed Mean (1,200) plus the average deviation (100), i.e. 1,300.

If the average deviation had been - 100 instead of + 100 then it would have been subtracted from the Gussed Mean instead of being added to it.

It is useful to note that should the actual Arithmetic Average have been taken as the Gussed Mean, then the deviations from it

would total zero. In the above case the actual Arithmetic Average being 1,300 then the deviations from it would be—

- 50
- 200
0
+ 50
+ 200

Total 0

Weighted Arithmetic Average

This is found by dividing the sum of the products of each item and its weight, by the sum of the weights. The "weight" may be the actual quantity for each item or it may be an estimated figure. In any case the significance of it is that it gives to each of the items the degree of importance which each should assume in the table.

The method of calculation of the Weighted Arithmetic Average will be seen from the following example.

Illustrative Example 3

It is desired to obtain the average wage paid per week per employee in a factory in which there are six different rates of labour. The following table sets out the position.

THE X.Y.Z. CO. LTD.		
Weekly Wages of Employees		
Week ended 8th February, 1950		
Wage per Week (£)	Number of Employees	Products
<i>x</i>	<i>w</i>	<i>w.x</i>
8	20	160
7	10	70
6	50	300
5	30	150
4	20	80
3	70	210
---	---	---
33	200	970
---	---	---

Here, the products are ascertained by multiplying the item (say £8) by its frequency (in this case 20) and so on. The Weighted Arithmetic Average is obtained by dividing the total of the products (970) by the total of the weights (200). This gives a Weighted Arithmetic Average of £4 17s. 0d., i.e. the average wage paid per week per employee is £4 17s. 0d.

It will be appreciated that it would have been misleading to have divided the total of the different wages (33) by their number (6) and claim that the average wage paid per employee was thus £5 10s. 0d. Such latter figure is merely an average of *rates*. If there had been only one employee receiving each of the different rates then £5 10s. 0d would have represented the average wage paid per employee as well as the average *rate* of wages paid in the factory.

To express the above by way of a formula:

- Let A.W. - The Weighted Arithmetic Average.
 x - The data expressed as individual items.
 w - The weights applicable to each item.

then

$$\text{A.W.} = \frac{\sum(wx)}{\sum(w)}$$

That is, the Weighted Arithmetic Average is equal to the sum of all the products obtained by multiplying each item by its weight, the whole divided by the sum of the weights.

Averaging where Class Intervals Used

In the case of a series wherein the items are shown in the form of class intervals (see page 21) the mid-point of each class interval may be taken as representing the particular class interval for the purpose of calculating the Arithmetic Average.

Thus if the class intervals are expressed as, say—

- *3— 6.99
- 7—10.99
- 11—14.99
- etc.

Then the mid-points are—

- 5
- 9
- 13
- etc.

If the frequencies applicable to each class interval were, say, 20, 50, 30 . . . etc., then the products would be 100, 450, 390 . . . etc.

Arithmetic Average and Trend

As intimated earlier, the Arithmetic Average does not indicate the direction in which a particular set of events may be moving. This is one of the dangers associated with Arithmetic Average which must be carefully guarded against. A comparison of the following tables will illustrate the position.

Company A			Company B		
Year		Net Profit £	Year		Net Profit £
1945	...	25,000	1945	...	47,000
1946	...	26,000	1946	...	38,500
1947	...	30,000	1947	...	35,000
1948	...	32,000	1948	...	32,000
1949	...	37,000	1949	...	30,500
1950	...	40,000	1950	...	29,000
1951	...	41,000	1951	...	19,000
		231,000			231,000

* Here the class interval is taken as 4. The upper limit (6.99, 10.99, etc.) in each case, has been taken short of the whole number above, so that the class intervals will not overlap. Thus 3—6.99 is to be taken as including 3 and every number up to but *not including* 7. The mid point is ascertained by adding half of the class interval (i.e. $\frac{4}{2}$) to the lower limit.

These are two series of Net Profits for Companies A and B over a period of seven years. It will be seen that the total profits are the same in each case and that therefore the Arithmetic Average of the Net Profit of each company is £33,000. Even a casual inspection of the two series, however, will reveal that whereas the profits of Company A are steadily increasing, those of Company B are steadily decreasing. If then we were simply told that the average profits of each company were £33,000, that information would not be a guide as to the progress of the Company in each case.

In respect of a company prospectus, promoters could supply misleading information by averaging profits in this way, if it were not for the conditions imposed by the 1948 Companies Act that the auditor or accountant must certify in the prospectus to *each* of the profits of the previous five years.

The question of *Trend* is dealt with in detail in a later chapter where certain other aspects of average such as "moving averages" and "semi-averages", etc. will be considered.

Advantages of the Arithmetic Average

The following is a summary of the advantages attaching to the Arithmetic Average as a particular means of averaging data:

- (1) It is the most widely used average and is generally more easily understood by people than are other averages.
- (2) It is comparatively simple to calculate.
- (3) The data does not have to be arranged in any particular order. Further, all that is required to calculate it is the total of the data and the number of items comprised in it.
- (4) It is used as the basis of further analysis of the series. This will be very much in evidence later, especially in the chapter dealing with Dispersion.
- (5) All items in the series are used in its calculation. In particular, where a weighted average is calculated, all items are given their relative share of importance in the calculation by means of the weights applicable to each of them.

Disadvantages of the Arithmetic Average

- (1) The Average may not be *truly* representative of the series in that the average as ascertained may not correspond with any particular item in the series. In illustrative Example 3 for instance, the average wage paid to employees was £4 17s. 0d. but no employee actually earned this wage. This of course may not always be a serious disadvantage.
- (2) Extreme items (i.e., very low or very high) may affect the average so that it will not be typical of the series.

- (3) The average tends to be dangerous if used on its own and without either the actual items themselves or an indication of trend. See under Arithmetic Average and Trend, *ante*, and under Dispersion, *post*.

THE MEDIAN

The *Median* is the value of the middle item in a series after it has been arrayed according to size.

To ascertain the Median, it is first necessary to arrange the series in order of size. The value of the middle item of the series so arrayed is the Median. Should the number of items in the series be an even number, then the Median may be taken as the average of the two middle items. If the number of items in the series is an odd number and it is desired to ascertain the median item it may be located by the formula

$$\frac{n + 1}{2}$$

where n represents the number of items in the series. Thus, if there are eleven items in the series then the median item would be

$$\frac{11 + 1}{2}, \text{ i.e. the 6th.}$$

Illustrative Example 4

Assume the following different wages have been paid to a group of employees and that it is desired to find the Median wage—

£ s. d.	£ s. d.	£ s. d.
4 2 6	3 7 6	6 4 6
7 - -	3 15 -	5 8 -
4 12 6	5 7 6	6 9 -
6 2 -	6 6 -	

Arrayed in order of size the above list of wages would appear as follows—

£ s. d.
3 7 6
3 15 -
4 2 6
4 12 6
5 7 6
5 8 -
6 2 -
6 4 6
6 6 -
6 9 -
7 - -

The Median wage, being the middle item, is therefore, £5 8s. 0d.

Another way of referring to the Median is to state that it is that item in the series such that there are as many items above it as there are below it. It is thus a measure of central tendency.

The illustration given above is rather simple and the slightly more difficult cases which may arise are—

- (a) Where the series contains both a characteristic *and* a frequency for each characteristic.
- (b) Where the series is arranged as in (a) but the characteristics are arranged in class intervals.

We will consider each of these cases in Illustrative Examples 5 and 6 respectively.

To bring out the comparison between the two workings, it should be noted that the respective series in these two examples are based on *exactly the same information*, the only difference between them being that they are *arranged differently*.

Illustrative Example 5

In the following distribution of employees by wages, the basic information derived from the wage cards is the wage (the characteristic) and the number of employees earning each particular wage (the frequency). Inserted in the table for purposes of calculation is the cumulative frequency. It is desired to find the Median Wage.

*Wage £ s. d.	Number of Employees	Cumulative Frequency
2 - -	20	20
2 5 -	30	50
2 10 -	40	90
2 15 -	10	100
3 - -	30	130
3 5 -	20	150
3 10 -	20	170
3 15 -	40	210
4 - -	20	230
4 5 -	10	240
4 10 -	50	290
4 15 -	60	350
5 - -	90	440
5 5 -	100	540
5 10 -	120	660
5 15 -	140	800
6 - -	150	950
	950	

The Median (being the mid-point) is situated at the point $\frac{950}{2}$ the 475th item, i.e., at £5 5s. 0d.

The Median Wage is therefore £5 5s. 0d.

It will be noted that the 475th item (proceeding down the Cumulative Frequency column) lies in the group with a frequency of 540 which represents therefore the Median Wage of £5 5s. 0d. since that is the wage at which the frequencies accumulate to 475.

* Note that in this series, the wages step up by 5s. in *contrast* to being divided into class intervals as set out in the next illustrative example.

Illustrative Example 6

Here, the same data as that applicable to Illustrative Example 5 has been arranged by class intervals and our purpose is to illustrate the finding of the Median where data is arranged in such manner.

£	*Wage		Number of Employees	Cumulative Frequencies
	£	s. d.		
2	2	15	100	100
3	3	15	110	210
4	4	15	140	350
5	5	15	450	800
6	6	15	150	950
			950	

We proceed by starting in the same way as for the previous example, i.e. since the total number of employees is 950 we divide that by 2 to find the mid-point, viz., 475.

The mid-point 475 lies this time in the group with a frequency of 800, i.e. somewhere between £5 and £5 15s. 0d. We may therefore call the class interval, £5 to £5 15s. 0d., the "Median group." To find the Median Wage we have now to find the *value* of the 475th item knowing that it lies somewhere between £5 and £5 15s. 0d. This is done by a procedure called "Interpolation".

The formula for interpolation is as follows—

$$M_e = L + \left(\frac{a \times CI}{b} \right)$$

- Where, M_e the Median.
 L the lower limit of the class interval in which the median item lies.
 a the difference between the median item and the number of items up to the lower limit of the class interval in which the median item lies.
 CI the class interval of the median group.
 b the frequency applicable to the class interval of the median group.

In this example then—

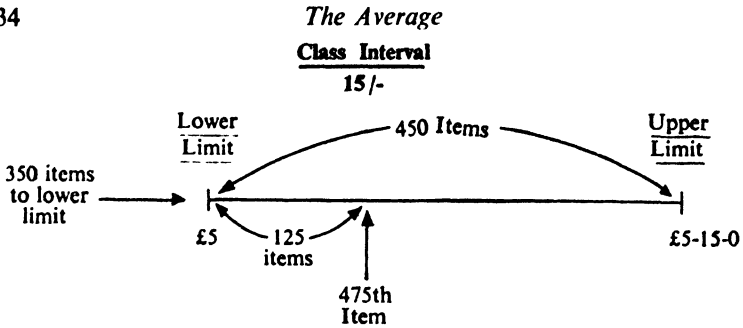
L	£5
a	125 (i.e., 475 less 350)
CI	£0 75 (i.e. 15s.).
b	450.

Therefore the Median is ascertained as follows—

$$\begin{aligned}
 M_e &= L + \left(\frac{a \times CI}{b} \right) \\
 &= \text{£}5 + \frac{125 \times 0.75}{450} \\
 &= \text{£}5 + 0.2083 \\
 &= \text{£}5.2083 \\
 &= \text{£}5 \text{ 4s. 2d.}
 \end{aligned}$$

The following diagram and explanation may help to explain the reasoning behind the formula.

* Note here that the class intervals are not *continuous*, i.e. the upper limit of one class does not correspond with the lower limit of the next class.



In this class interval are 450 items. To find how far into this class interval the item 475 extends, we take from 475 the number of items up to the lower limit (i.e. 350 items) the result being 125. So then the *value* of the Median item of 475 is the lower limit (5) plus the value of the class interval (15s.), multiplied by the proportion of the number of items extending into the class interval (125) to the total number of items in the class interval (450).

Comparison of Methods of Finding the Median

The method of ascertaining the Median described in Illustrative Examples 4 and 5 may be termed the Direct Method. That is, virtually what is done is that all the items are listed down in order of size and the centre one is picked and thus the value attaching to that centre item is the Median. In Illustrative Example 5 it will be seen that had all of the 950 items been listed one after the other from £2 to £6 then the centre or 475th item would have been £5 5s. 0d.

In Illustrative Example 6 we took the same data as that in 5 but re-arranged it in class intervals and found the Median to be £5 4s. 2d. The reason for taking the same data was to show how closely the two methods approximated in result. The difference in the latter case is caused by the fact that we only knew that the Median lay somewhere between £5 and £5 15s. 0d. and had to find a particular figure by interpolation, i.e. on a proportional basis. It was assumed in Illustrative Example 6 that the table there presented was the only information we had and thus we did not know *how* the different items were spread throughout each class interval. If we had known that, we would have used the direct method illustrated in Example 5. But where the detail is thus not known then the method of interpolation is the only one that can be used.

It is interesting to note that the *Arithmetic Average* of the series as set out in Illustrative Example 5 is £4 16s. 6d.

Advantages of the Median.

The following is a summary of the advantages of the use of the Median as a particular form of average:

- (1) It is easily calculated. In many cases it may be obtained by mere inspection of the series.

- (2) It is not affected by extreme items in the series. This is because its location depends not on the magnitude of the items but on the number of them.
- (3) It is typical of the series in that it is an actual item (that is, if found directly. If found by interpolation it may not in some cases be an actual item. Again in some cases it may be the average of the two centre items.
- (4) It is easily understood.

Disadvantages of the Median

- (1) Before the Median can be located, the data must be arrayed in order of size. This may present some difficulty where the data is voluminous.
- (2) Mathematically, it is of limited use, e.g. if the Medians of sub-groups were ascertained, then such Medians could not be averaged to find the Median of the whole group.
- (3) It is not likely to be representative of a series in which there are only a few items.

THE MODE

The Mode (or as it is sometimes termed, the Norm) is another measure of central tendency and is the position of greatest numerical density. It is really the most common value, i.e. the value which occurs the greatest number of times.

In the following series the Modal Average of products sold is Product E, since there were more of this product sold than any other.

Product	Sales £
A	92
B	46
C	78
D	20
E	150
F	75
G	43
H	89

In the above case the Mode was obtained by simple inspection, but in many instances it is more difficult to ascertain due to the fact that the data may be more evenly spread and there may be no apparent item which occurs more times than any other or at least it may not be obvious that there is a point of greatest density in the series. In such a case a method of arriving at the point of greatest density in the series is that of continued grouping as illustrated in the following example.

Illustrative Example 7

The following is the result of a survey taken of the length of time each employee takes to carry out a certain task. It was found that

the shortest time taken was 5 minutes and that the longest was 12 minutes. It is desired to find the Modal time.

Time taken (Minutes)	Number of Employees
5	22
6	6
7	17
8	30
9	10
10	9
11	30
12	4

The first step is to group the frequencies successively in two's and three's in order to see which group predominates.

Time	Number	(a)	(b)	(c)	(d)	(e)
5	22	28	23	45	53	57
6	6					
7	17	47				
8	30		40			
9	10	19		49		
10	9					
11	30	34	39		49	43
12	4					

The columns (a) to (e) represent the five groupings of the data. Columns (a) and (b) are groupings in pairs and columns (c), (d) and (e) are groupings in three's. From these groupings the following observations are made—

Column (a), largest group is 47 and contains Times 7 and 8
 .. (b), 40 8 .. 9
 .. (c), 49 8 .. 9 and 10
 .. (d), 53 6 .. 7 .. 8
 .. (e), 57 7 .. 8 .. 9

It can thus be seen that Time 8 with its frequency of 30 employees is the only one common to all the largest of the groupings.

Time of 8 minutes is therefore the Modal Time.

Actually it may not always be necessary to carry out the groupings as far as was done in this instance. If the above example is again referred to it will be seen that it is fairly apparent that from column (a) alone, Time 8 is the measure of central tendency. However, the groupings were carried further for the purpose of illustrating how they would be continued if that were not so after the first grouping.

Series with Class Intervals

Where the characteristic of the series is arranged in class intervals the Mode must be found by means of a formula.

The formula is as follows:

$$M_o = L + \frac{a \times CI}{a + b}$$

- where, M_o -- The Mode.
 L -- The lower limit of the modal group.
 a -- The frequency or number of items in the next higher class.
 b -- The frequency or number of items in the next lower class.
 CI -- The class interval.

Illustrative Example 8

It is desired to find the Modal Wage from the following series (which incidentally is the same as that in Illustrative Example 6).

Wage				Number of Employees
£	£	s.	d.	
2	---	2	15	100
3	---	3	15	110
4	---	4	15	140
5	---	5	15	450
6	---	6	15	150

By inspection the Modal Wage lies in the class interval £5-£5 15s. 0d. which class interval represents the modal group.

Applying the formula we have—

$$M_o = L + \frac{a \times CI}{a + b}$$

$$= \text{£}5 + \frac{150 \times 0.75}{150 + 140}$$

$$= \text{£}5 + 0.387$$

Mode = £5 7s. 9d.

Advantages of the Mode

The following is a summary of the advantages attaching to the use of the Mode as a form of average:

- (1) In many cases it can be obtained by inspection of the data.
- (2) Since it is the value which has occurred the greatest number of times, it is the most descriptive average.
- (3) It is generally quite easily understood.
- (4) It is not affected by extreme items.
- (5) If further examples are collected the Mode is the one most likely to occur.

Disadvantages of the Mode

- (1) Where the number of items is small the Mode may not be representative and further, if the class intervals are large it may be difficult to determine.
- (2) The frequency distribution to which it relates must be comparatively large and possess a distinct central tendency. In fact in a particular distribution it may not exist because of the fact that none of the values may be repeated.
- (3) Once ascertained it is not capable of further algebraic manipulation.

THE GEOMETRIC MEAN

If there are n terms in a series, then the Geometric Mean (Symbol M_g) of the series is the n th root of the product of such n terms.

The following illustrations show how it is computed—

Series	Geometric Mean
(a) 4, 9	$\sqrt{4 \times 9} = 6$
(b) 1, 1, 27	$\sqrt[3]{1 \times 1 \times 27} = 3$
(c) 1, 4, 4, 16	$\sqrt[4]{1 \times 4 \times 4 \times 16} = 4$

It will be noted that in series (a) there are 2 terms, in series (b) there are 3 terms and in series (c) there are 4 terms. Even though the above series do not mean very much it will be noted that the effect of the Geometric Mean is to weight down the large items in the series and weight up the small items. That is the general effect of the Geometric Mean.

Illustrative Example 9

Compute the Geometric Mean of the following series.

Year	Indexes
1948	110
1949	124
1950	105
1951	109
1952	130

$$\begin{aligned} \text{Geometric Mean} &= \sqrt[5]{110 \times 124 \times 105 \times 109 \times 130} \\ \therefore \text{Log G.M.} &= \frac{1}{5} (\log 110 + \log 124 + \log 105 + \log 109 + \log 130) \\ &= \frac{1}{5} (2.0414 + 2.0934 + 2.0212 + 2.0374 + 2.1139) \\ &= 2.0614 \end{aligned}$$

$$\therefore \text{Geometric Mean} = \text{anti-log of } 2.0614 = 115.2.$$

Tables of Logarithms are given in the Appendix at the end of this Volume.

Geometric Mean and Price Indexes

Since the Geometric Mean measures the relative difference between the items (or the ratio of change) it is particularly applicable in the computation of Index Numbers. The following example illustrates how the Geometric Mean brings to light the true state of affairs in a case which would be falsely represented by the Arithmetic Average.

Illustrative Example 10

The following table sets out the prices attaching to four different products for each of the years 1948, 1949.

Product	1948		1949	
	£	s. d.	£	s. d.
A	—	15 —	—	7 6
B	2	6 —	4	12 —
C	—	14 —	1	8 —
D	10	— —	5	— —

These could be represented by indexes as follows, assuming the base year is 1948 and that the price in such base year is 100.

Product	1948	1949
A	100	50
B	100	200
C	100	200
D	100	50
	<hr style="width: 50px; margin: 0 auto;"/>	<hr style="width: 50px; margin: 0 auto;"/>
	400	500
	<hr style="width: 50px; margin: 0 auto;"/>	<hr style="width: 50px; margin: 0 auto;"/>
Arithmetic Average	100	125

Now if we were to assume from the above that there was an increase in the general level of prices of these four products of 25 per cent. we would be wrong. That, however, is what the Arithmetic Average would lead us to believe.

Actually if we examine the indexes we will see that the general condition of the prices is the same in 1949 as that in 1948. The calculation of the Geometric Mean will prove this.

The Geometric Mean for 1948

$$= \sqrt[4]{100 \times 100 \times 100 \times 100}$$

$$= 100.$$

The Geometric Mean for 1949

$$= \sqrt[4]{50 \times 200 \times 200 \times 50}$$

$$= 100.$$

This is so since the prices of two products have doubled and the prices of the other two have halved, leaving the general price condition the same. This particular example is only given to illustrate the *theory*, and would not represent the position in practice where different weights would attach to each of the items—unlike this case where it is assumed that all the items have equal weights.

Advantages of the Geometric Mean

The following are the advantages of the use of the Geometric Mean as a particular form of average.

- (1) It weights down large items in the series and weights up small items.
- (2) It uses all of the data in the series.
- (3) The influence of large items is not reflected in the result so much as is the case with the Arithmetic Average.
- (4) It is not necessary to arrange the data in any special way.
- (5) If the Geometric Average and the number of items are given the aggregate can be found.

Disadvantages of the Geometric Mean

- (1) It cannot be used where one of the items in the series is zero, or a minus quantity.
- (2) It is not generally understood as easily as are the Arithmetic Average, the Median and the Mode.

- (3) It is more involved in its calculation than are the types of averages mentioned in (2).
- (4) The result may not be representative of any particular item in the series.

THE HARMONIC MEAN

The Harmonic Mean is used in the averaging of *rates*.

If the items in the series are designated by the symbols— $X_1, X_2, X_3, X_4, X_5 \dots X_n$ (there being n of them) then the Harmonic Mean is computed from the formula—

$$M_h = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{1}{x_5} + \dots + \frac{1}{x_n}}$$

that is, it is the reciprocal of the arithmetic average of the reciprocals of the values.

Illustrative Example 11

Calculate the Harmonic Mean of the series—

2, 4, 5, 10, 20.

Solution:

$$\begin{aligned} M_h &= \frac{1}{\frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{10} + \frac{1}{20}} \\ &= \frac{10 + 5 + 4 + 2 + 1}{20 \times 5} \\ &= \frac{22}{100} \\ \therefore M_h &= \frac{100}{22} = 4.54 \end{aligned}$$

Advantages of the Harmonic Mean

The following is a summary of the advantages attaching to the use of the Harmonic Mean as a particular form of average.

- (1) It is particularly adapted for averaging rates.
- (2) It is capable of algebraic manipulation.
- (3) It gives little weight to extreme items.

Disadvantages of the Harmonic Mean

- (1) It is not generally as well understood as the Arithmetic Average, the Median or the Mode.
- (2) Its calculation is more involved than are those for the averages mentioned in (1).

CHAPTER 3

DISPERSION

INTRODUCTION	41
MEASURES OF DISPERSION			43
THE RANGE	43
THE QUARTILE DEVIATION	44
Illustrative Example 1.									
Illustrative Example 2.									
Quartile Coefficient of Dispersion.									
THE AVERAGE DEVIATION	47
Illustrative Example 3.									
Illustrative Example 4.									
THE STANDARD DEVIATION	50
Illustrative Example 5.									
Illustrative Example 6.									
COEFFICIENTS OF DISPERSION					53
Illustrative Example 7.									
SKEWNESS	55

INTRODUCTION

In the previous two chapters, consideration was given to the fundamentals of—

- (i) the statistical series itself, and
- (ii) the average.

However, whilst use may be made of the series and the average as they stand, in the majority of cases, these can only be considered as first steps in the analysis of the data. What are generally required next are answers to the following questions:

- (a) Is the average really typical of the series?
- (b) To what extent do the various items comprising the data vary from the average?
- (c) What is the spread of the data?

Dispersion is a measure which tells these things and the methods of measuring dispersion take different forms according to what is required. Dispersion, defined simply, is a measure of the scatter or variation of the items in a series, about the average.

The importance to be attached to the question stated in (a) above, cannot be over-emphasized. Much confusion and many errors result from the false idea assumed by some that, because an average of a series represents that series, then you can compare different series by comparing their respective averages. To illustrate the danger associated with such a concept, assume that we have two factories, the Alpha factory and the Beta factory, and that the former manufactures soap and the latter furniture. Their respective outputs are recorded in value for each month of the year 1950 as follows:

Month	Alpha Factory	Beta Factory
	Value of Production	Value of Production
	£	£
January	9,000	2,900
February	2,000	2,700
March	Nil	3,000
April	3,000	3,200
May	1,000	3,100
June	2,000	2,800
July	3,000	3,100
August	Nil	3,200
September	5,000	2,950
October	6,000	2,850
November	2,000	3,200
December	3,000	3,000
	<hr/>	<hr/>
	£36,000	£36,000
	<hr/>	<hr/>
Monthly Average (Arithmetic Average)	£3,000	£3,000

Now, it is clear that by simply inspecting the averages of each of these two series, we cannot judge of their relative nature, i.e. we cannot say that the Alpha Factory has a satisfactory output at £3,000 per month *because* the Beta Factory has a similar value of output. Nor could we claim contrariwise, i.e. that the output of the Alpha Factory was unsatisfactory because it was *only* equal to that of the Beta Factory. We would have no right to make such assumptions from a mere inspection of each of the averages.

Looking at each of the series it is obvious that whilst the average of the figures of the Beta Factory *is* a true and faithful representation of that series as a whole, the same could *not* be said of the figures as disclosed by the Alpha Factory. *We can only make this statement, however, after inspecting the data from which each of the averages was derived and seeing what the degree of representation actually was.* If we had not made such an inspection we would not have known, *in either series*, whether the average was representative or not. Judging by the average above, we could have made no assumption, *in either series*, about the output conditions in either factory.

Summarizing these observations we have the following points:

- (a) In each case (i.e. in each series) we do not know whether the respective average of each is typical of its series unless we inspect the data from which the averages were derived.

- (b) A comparison of the two series by means of their averages thus cannot be made, quite apart from the fact that in this particular case the outputs of each factory are of different commodities.

In the above illustration, two series were considered simply for the purpose of emphasis. Our remarks apply with equal force to *any particular* series and it is clear from the explanation given, that in addition to the average of a series, we require some further information. The measures of dispersion outlined below are devices designed to supply this need and demonstrate the spread of the items in a series, or the deviation of the individual items from the average of the series.

MEASURES OF DISPERSION

The different measures of dispersion which are commonly used are four in number and may be conveniently grouped as follows:

- A. Those which measure the spread of the data without taking into account the relation of the average to the series. They are:
1. The Range.
 2. The Quartile Deviation.
- B. Those which ascertain how far all the items of the series vary from the average. They are:
3. The Average Deviation.
 4. The Standard Deviation.

1. THE RANGE

The *Range* is simply an expression of the difference between the maximum and minimum values of the items in the series. In the following series the *Range* is £1 6s. 0d.

Product Type	Retail Price		
	£	s.	d.
A	2	-	-
B	2	5	-
C	2	7	6
D	3	-	-
E	3	6	-

The *Range*, as used here, has no technical meaning other than that supplied by common usage, so that if a trader states that the retail prices of his products *range* from £2 to £3 6s. 0d., he is only indicating the lower and upper limits of the prices between which his prices of all goods are spread. As a measure of dispersion, the *Range* is of limited use, particularly in the case of series wherein the extremes are of an exceptional nature. For instance, if a trader deals in goods all of which lie in price between £2 and £3, except

that in addition he sells gramophone needles at 1s. a tin and refrigerators at £100 each, it is little to the point to state that the Range of his retail prices is £99 19s. 0d. The exceptionally low extreme and the exceptionally high extreme have, in this case, distorted the position.

In some cases the Range is extremely useful to Management, e.g. where it is a pointer to the evenness of output over a period. In the figures of the Beta Factory shown on page 42 the lowest value is £2,700 and the highest £3,200, the Range thus being £500. It is obvious that the more this value of the Range can be reduced, the more constant and steady will be the output, month by month. If this value were taken as standard, say for this year, then a drop in value to say £300 next year would indicate a still more satisfactory state as far as constant production was concerned, assuming that prices of materials, etc. have remained steady.

2. THE QUARTILE DEVIATION

This is the other measure of Dispersion which measures the spread of the data without taking into account the relation of the average to the series.

The *Quartiles* are the values which are *one-quarter* and *three-quarters* of the way up a series of items, after such series has been arrayed in order of size. It will be remembered that the *middle* value of an array is termed the Median, so that we can say that the Median, together with the Quartiles, divide the series into four equal parts.*

The *Quartile Deviation* is found from the following simple formula—

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

where, Q.D. — the Quartile Deviation.
 Q_3 — the upper Quartile.
 Q_1 — the lower Quartile.

It will be evident that Q_3 and Q_1 between them enclose half the area of the series—furthermore, it is the central area which is so covered and thus the Quartile Deviation takes no account of extremes, as does the Range. Sometimes the Quartile Deviation is called the Semi Inter-Quartile Range.

Illustrative Example 1

From the following list of the costs of production of different units produced in a factory find the Quartile Deviation. These items are arrayed in order of size.

Cost Per Unit
£
1.0
1.2

* Other measures which correspond to the Quartiles are—
 Quintiles, dividing the series into 5 equal parts
 Deciles, " " " " 10 " "
 Percentiles " " " " 100 " "

1.4	1st Quartile (Q ₁)
1.8		
2.0		
2.2	Median
2.6		
2.8		
3.0	3rd Quartile (Q ₃)
3.2		
3.4		

$$\begin{aligned} \therefore \text{Quartile Deviation} &= \frac{Q_3 - Q_1}{2} \\ &= \frac{3.0 - 1.4}{2} \\ &= \text{£}0.8 \end{aligned}$$

Should the series have consisted of 9 items then the Quartiles would have to be found by averaging as follows. Assume the same series, omitting the last two items.

Cost Per Unit		
£		
1.0		
1.2	}1st Quartile = $\frac{2.6}{2} = 1.3$
1.4		
1.8		
2.0	Median
2.2		
2.6	}3rd Quartile = $\frac{5.4}{2} = 2.7$
2.8		
3.0		

Thus the 1st Quartile is obtained by averaging the 2nd and 3rd items and the 3rd Quartile by averaging the 7th and 8th items.

The Quartile Deviation here would be—

$$\begin{aligned} \text{Q.D.} &= \frac{Q_3 - Q_1}{2} \\ &= \frac{2.7 - 1.3}{2} \\ &= \text{£}0.7 \end{aligned}$$

Illustrative Example 2

The following is a distribution of employees by wages. Calculate the Quartile Deviation.

Wage Per Week						No. of Employees	Cumulative Frequency
£	s.	d.	£	s.	d.		
2	-	-	2	9	11	...	214
2	10	-	2	19	11	...	326
3	-	-	3	9	11	...	92
3	10	-	3	19	11	...	46
4	-	-	4	9	11	...	192
4	10	-	4	19	11	...*	78
5	-	-	5	9	11	...	30
5	10	-	5	19	11	...	49
6	-	-	6	9	11	...	90
6	10	-	6	19	11	...	151
7	-	-	7	9	11	...	76
						1,344	

It is first necessary for us to locate the 1st and 3rd Quartiles.

The 1st Quartile is located at the point $\frac{1344}{1} \times \frac{1}{4}$ i.e. 336, and is therefore in the group £2 10s. 0d. to £2 19s. 11d.

The 3rd Quartile is located at the point $\frac{1344}{1} \times \frac{3}{4}$, i.e. 1008 and is therefore in the group, £5 10s. 0d. to £5 19s. 11d.

We have located the Quartiles and must next proceed to find their values, which should be done by the process of interpolation as was necessary in evaluating the median where it was known that it was contained in a certain class interval (see page 33).

1st Quartile.

$$\begin{aligned} \text{* By interpolation } Q_1 &= L + \left(\frac{a \times CI}{b} \right) \\ &= \text{£2 10s. 0d.} + \left(\frac{122 \times 0.5}{326} \right) \\ &= \text{£2 10s. 0d.} + 0.187 \\ &= \text{£2 10s. 0d.} + \text{£0 3s. 9d.} \\ &= \text{£2 13s. 9d.} \end{aligned}$$

3rd Quartile.

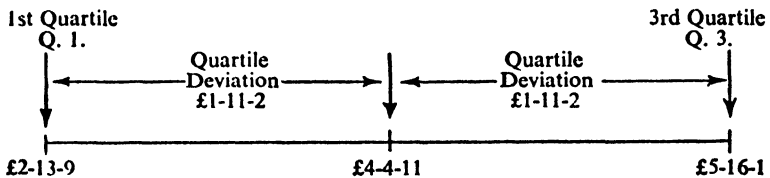
$$\begin{aligned} \text{By interpolation } Q_3 &= L + \left(\frac{a \times CI}{b} \right) \\ &= \text{£5 10s. 0d.} + \left(\frac{30 \times 0.5}{49} \right) \\ &= \text{£5 10s. 0d.} + 0.306 \\ &= \text{£5 10s. 0d.} + \text{£0 6s. 1d.} \\ &= \text{£5 16s. 1d.} \end{aligned}$$

The Quartile Deviation is therefore—

$$\begin{aligned} \frac{Q_3 - Q_1}{2} &= \frac{\text{£5 16s. 1d.} - \text{£2 13s. 9d.}}{2} \\ &= \frac{\text{£3 2s. 4c.}}{2} \\ &= \text{£1 11s. 2d.} \end{aligned}$$

We can now see what the expression, “Quartile Deviation”, means. It is clear that by taking the 1st and 3rd Quartiles, we consider or work upon *half* of the total items in the distribution (since 1008, the 3rd Quartile, less 336, the 1st Quartile = 672, i.e. half of the total items of 1344).

Therefore, viewing this half of the distribution, it would appear in diagrammatic form as follows:



- * L = the lower limit of the class interval in which the quartile lies.
- a = the difference between the quartile item and the number of items up to the lower limit of the class interval in which the quartile item lies.
- CI = the class interval of the quartile group.
- b = the frequency applicable to the class interval of the quartile group.

A useful way of describing this state of affairs is to say that one half of all the employees receive wages falling within the range £4 4s. 11d. \pm £1 11s. 2d. It should be noted that the value of £4 4s. 11d. (calculated by adding the quartile deviation to the 1st Quartile) will only coincide with the median provided the distribution is perfectly symmetrical.

Quartile Coefficient of Dispersion

The Quartile Deviation which we have just considered is an *absolute* measure and thus is not always suitable for using for comparative purposes. That is, if we desire to compare two different series as regards their quartile deviations the actual or absolute measures applicable to each are not comparable.

To enable us to make the comparison we must first convert the quartile deviation to a *coefficient*, and when so converted it is known by the title, Quartile Coefficient of Dispersion (note that the Quartile Deviation is one of the methods of Dispersion). The meaning of a coefficient of dispersion is that it is a *relative* measure as contrasted with measures of dispersion which are *absolute*. It may happen for instance that two series, say A and B, have the same measure of dispersion but different averages. These two are therefore not comparable by their measures of dispersion because if, say, the *average* of Series A were larger or greater than that of Series B then there is relatively less variation in Series A than in Series B.

As pointed out previously, the range and the quartile deviation are measures of dispersion which do not take into account the average of the series and therefore there is no coefficient of dispersion applicable to the range. In the case of the Quartile Deviation, the Quartile Coefficient of Dispersion is calculated according to the following formula—

$$\text{Q.C.D.} = \left(\frac{Q_3 - Q_1}{\frac{Q_3 + Q_1}{2}} \right), \text{ i.e., } \frac{Q_3 - Q_1}{\frac{Q_3 + Q_1}{2}}$$

where Q.C.D. = the Quartile Coefficient of Dispersion.

In illustrative example 2 the Quartile Coefficient of Dispersion would be—

$$\begin{aligned} \text{Q.C.D.} &= \frac{\text{£5 } 16\text{s. } 1\text{d.} - \text{£2 } 13\text{s. } 9\text{d.}}{\text{£5 } 16\text{s. } 1\text{d.} + \text{£2 } 13\text{s. } 9\text{d.}} \\ &= \frac{5.80 - 2.68}{5.80 + 2.68} \\ &= \frac{3.12}{8.48} \\ &= 0.37 \text{ or } 37 \text{ per cent.} \end{aligned}$$

3. THE AVERAGE DEVIATION

As indicated earlier, the Average Deviation is a measure of Dispersion which aims to demonstrate how far the individual items

of the series vary from the average. The term "average" used here may mean the Arithmetic Average, Median or Mode, but since the Arithmetic Average is generally used as the central point from which deviations are measured, this will be used in illustration.

In arriving at the Average Deviation the following procedure is carried out—

- (i) Calculate the Arithmetic Average of the original data.
- (ii) Ascertain the deviation of each item from this average.
- (iii) Add the total of the deviations (ignoring the algebraic signs) found from (ii).
- (iv) Divide this total by the number of items (*not* by the frequencies) and the result is the Average Deviation.

Illustrative Example 3

The following frequency distribution represents the outputs of seven employees of a factory for a particular month. All are engaged on the same type of task and the output is expressed in pounds weight. Calculate the average deviation.

Employee	Output in pounds
1	70
2	53
3	49
4	81
5	67
6	58
7	49

Solution:

Employee	Output in Pounds	Deviation from Arithmetic Average
1	70	9
2	53	8
3	49	12
4	81	20
5	67	6
6	58	3
7	49	12
	<hr style="width: 50px; margin: 0 auto;"/> 427	<hr style="width: 50px; margin: 0 auto;"/> 70
Arithmetic Average =	$\frac{427}{7} = 61$ lb.	
Average Deviation =	$\frac{70}{7} = 10$ lb.	

In the following illustrative example the Average Deviation is computed in respect of two series for comparative purposes.

Illustrative Example 4

The following series represent the sales of two departments for the period indicated. Calculate the Average Deviation of each of the departmental sales.

							Dept. A	Dept. B
							£	£
February	2	32	53
	3	26	12
	4	30	20
	5	22	31
	6	45	49
							<u>155</u>	<u>165</u>

Solution:

	DEPT. A			DEPT. B	
	Sales	Deviation from	Sales	Deviation from	
	£	Arith. Average	£	Arith. Average	
February	2	32	1	53	20
	3	26	5	12	21
	4	30	1	20	13
	5	22	9	31	2
	6	45	14	49	16
		<u>155</u>	<u>30</u>	<u>165</u>	<u>72</u>
Arithmetic					
Average					
		31		33	
Average					
Deviation			$\frac{30}{5} = £6$		$\frac{72}{5} = £14.4$

In the calculation set out above it will be noted that when the deviation from the average is ascertained, no account is taken of the *sign* of the resultant, i.e. whether the resultant is positive or negative. This is so because algebraically the sum of the deviations from the average would total zero and thus no further calculation would be possible. To avoid this the algebraic sign is ignored, an admitted weakness in the procedure which is eliminated in the method used in the calculation of the Standard Deviation, described later.

The working illustrated in the example may be expressed by way of a formula as follows:

$$\delta = \frac{\Sigma d}{n}$$

Where δ the average deviation.
 d the deviation of the value of each of the items from the average (the sign Σ means "the sum of").
 n the number of items.

This deviation when calculated is an *absolute* measure of dispersion and to obtain the coefficient of dispersion, or the relative measure, it is necessary to divide the average deviation by the average.

In Illustrative Example 4 the coefficient of dispersion would be calculated as follows—

						Dept. A	Dept. B
Average	31	33
Average Deviation	<u>6</u>	<u>14.4</u>
Coefficient of Dispersion		$\frac{6}{31}$	$\frac{14.4}{33}$
						£0.19 or 19%	£0.43 or 43%

The significance of these comparative coefficients is that since the fluctuation about the average sales is far greater in Dept. B than in Dept. A, the sales for the latter are steadier and more constant than the former and perhaps it may mean too that they are more reliable and more certain. In practice, however, comparison over a longer period than is covered in this example would be desirable before forming conclusions.

4. THE STANDARD DEVIATION

This measure of dispersion is similar to the average deviation, except that the deviations when ascertained are squared and the square root of the average of the total of these squares is taken as the Standard Deviation. The standard deviation is capable of being treated algebraically and the steps necessary in its calculation may be summarized as follows:

- (i) Calculate the arithmetic average of the original data.
- (ii) Ascertain the deviation of each item from this average.
- (iii) Square each of the deviations so obtained.
- (iv) Add the squares of the deviations.
- (v) Divide the sum of the squares by the number of items in the original data.
- (vi) Take the square root of the result of (v) and this represents the standard deviation.

In the following example the same data as in Illustrative Example 4 is used to indicate the similarity in the two methods.

Illustrative Example 5

	DEPT. A			DEPT. B		
	Sales	Deviations (d)	Deviations Squared (d ²)	Sales	Deviations (d)	Deviations Squared (d ²)
Feb. 2	32	1	1	53	20	400
3	26	5	25	12	21	441
4	30	1	1	20	13	169
5	22	9	81	31	2	4
6	45	14	196	49	16	256
	<hr/> 155		<hr/> 304	<hr/> 165		<hr/> 1,270
Arithmetic Average	<hr/> 31			<hr/> 33		

Now, the Standard Deviation of Dept. A is the square root of $\frac{304}{5}$, i.e.,

$$\sqrt{\frac{304}{5}} \text{ or } 7.79$$

and the Standard Deviation of Dept. B is the square root of $\frac{1270}{5}$, i.e.

$$\sqrt{\frac{1270}{5}} \text{ or } 15.93$$

(The arithmetical method of calculating the square root is explained in the footnote.)*

The use of the Standard Deviation is preferred to that of the Average Deviation because it is more correct mathematically in its calculation, in that the *sign* of the resultant of the subtraction of the item from its average (e.g. 32 from 31 in Illustrative Example 4) is ignored when computing the Average Deviation. If it were not so ignored, the sum of the deviations would be zero and so prevent further analysis. With the Standard Deviation, however, all the

* The ordinary arithmetical method of calculating the square root of a number is as follows:

Take a number, say, 72,467.

Pair off the numerals from the right, thus—

7,24,67.

Now find a number such that when multiplied by itself gives a number as near as possible to 7. This would be 2. Then we have—

$$\begin{array}{r} 2 \\ \hline 7,24,67 \\ 4 \\ \hline 3 \end{array}$$

Then bring down the next pair, and find a number (x) such that when placed after 2 doubled (i.e. 4) (resulting in forty-x) and multiplied by x, gives a product as close as possible to 324. This would be 6 so that the calculation would appear as follows—

$$\begin{array}{r} 26 \\ \hline 7,24,67 \\ 4 \\ \hline 46 \quad 324 \\ 276 \\ \hline 48 \end{array}$$

This process is continued and when the last pair of numerals is disposed of, pairs of noughts may be brought down if the square root is required to one or several places of decimals. The remainder of the above calculation would be as under—

$$\begin{array}{r} 269.19 \\ \hline 7,24,67 \\ 4 \\ \hline 46 \quad 324 \\ 276 \\ \hline 529 \quad 4867 \\ 4761 \\ \hline 5381 \quad 10600 \\ 5381 \\ \hline 53829 \quad 521900 \\ 484461 \\ \hline 37439 \end{array}$$

So, the square root of 72,467 is 269.19 to two places of decimals.

deviations from the average are squared and it thus makes no difference whether such deviations were positive or negative quantities.

The formula for the Standard Deviation is as follows:

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

Where σ - the standard deviation.
 d^2 - the squares of the deviations of the items from the average.
 n - the number of items.

As indicated previously, the dispersions of different series cannot be compared by aligning their respective *measures* of dispersion, which are *absolute* values. It is necessary to calculate the coefficient of dispersion which in the case of the standard deviation is found by dividing the standard deviation by the arithmetic average. In the example above, the coefficient of dispersion would be ascertained as follows:

		Dept. A	Dept. B
Standard Deviation	7 79	15 93
Arithmetic Average	31 00	33 00
		<hr style="width: 50%; margin: 0 auto;"/>	<hr style="width: 50%; margin: 0 auto;"/>
Coefficient of Dispersion	7.79	15.93
		<hr style="width: 50%; margin: 0 auto;"/>	<hr style="width: 50%; margin: 0 auto;"/>
		31	33
		0.25	0.48
		or 25%	or 48%

When comparing, say, the output of two employees in a factory, it is often very useful not only to know what is the average output of each but also how far or to what extent such output fluctuates about his average. Assuming that in Illustrative Example 5, Depts. A and B represent two employees A and B and that the figures in the Sales column represent the production in pounds (weight) of each, then the coefficient of dispersion of A being 25 per cent. and that of B 48 per cent., it is clear that A would be a much steadier and more reliable worker than B, because his production figures fluctuate far less about the average than do these of B. This, of course, is so provided that A and B are producing similar articles.

The following illustrative example shows the method of calculating the standard deviation where, in the series, the items are shown in the form of class intervals.

Illustrative Example 6

Compute the standard deviation of the following frequency distribution.

Exceeding	Not Exceeding	Frequency
5.5	6.5	10
6.5	7.5	18
7.5	8.5	9
8.5	9.5	13

Solution:

It is first necessary to calculate the arithmetic average and this is done as follows:

Exceeding	Not Exceeding	Mid Point	Frequency	Product of Mid Point and Frequency
5.5	6.5	6	10	60
6.5	7.5	7	18	126
7.5	8.5	8	9	72
8.5	9.5	9	13	117
			50	375

$$\begin{aligned} \text{Arithmetic Average} &= \frac{375}{50} \\ &= 7.5 \end{aligned}$$

The standard deviation is then calculated as follows:

Exceeding	Not Exceeding	Mid Point	Frequency	Deviations	Deviations Squared	Frequency, times Deviations Squared
5.5	6.5	6	10	1.5	2.25	22.50
6.5	7.5	7	18	0.5	0.25	4.50
7.5	8.5	8	9	0.5	0.25	2.25
8.5	9.5	9	13	1.5	2.25	29.25
			50			58.50

$$\begin{aligned} \text{Standard Deviation} &= \sqrt{\frac{58.50}{50}} \\ &= 1.08 \end{aligned}$$

COEFFICIENTS OF DISPERSION

Although each of the various coefficients of dispersion was considered when each of the respective measures of dispersion was dealt with, it is thought convenient to repeat them here in summary form so that by collecting them together under the one head, emphasis is made as to when each is used.

Each of the three measures of dispersion, viz.—

- (1) The Quartile Deviation.
- (2) The Average Deviation.
- (3) The Standard Deviation

has its coefficient of dispersion which, in formula form, is respectively—

- (1) $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ where $\begin{cases} Q_1 = \text{1st Quartile} \\ Q_3 = \text{3rd Quartile} \end{cases}$
- (2) $\frac{\delta}{A}$ where $\begin{cases} \delta = \text{Average Deviation} \\ A = \text{Average} \end{cases}$
- (3) $\frac{\sigma}{A}$ where $\begin{cases} \sigma = \text{Standard Deviation} \\ A = \text{Average} \end{cases}$

Illustrative Example 7

Compare the dispersions of the following series, making the necessary calculations based on the Standard Deviation.

Year	Company A Profits in £s	Company B Profits in £s
1	3,600	1,020
2	5,400	1,140
3	4,800	1,080
4	5,200	1,300
5	6,300	1,400
6	6,500	1,500
7	7,200	1,800
8	6,000	1,450
9	5,400	1,010

Solution:

Year	COMPANY A			COMPANY B		
	Profit	Deviation from Arithmetic Average (5,600)	Deviations squared	Profit	Deviation from Arithmetic Average (1,300)	Deviations squared
	£			£		
1	3,600	-2,000	4,000,000	1,020	-280	78,400
2	5,400	-- 200	40,000	1,140	-160	25,600
3	4,800	-- 800	640,000	1,080	-220	48,400
4	5,200	-- 400	160,000	1,300	—	—
5	6,300	700	490,000	1,400	100	10,000
6	6,500	900	810,000	1,500	200	40,000
7	7,200	1,600	2,560,000	1,800	500	250,000
8	6,000	400	160,000	1,450	150	22,500
9	5,400	-- 200	40,000	1,010	-290	84,100
	<u>50,400</u>		<u>8,900,000</u>	<u>11,700</u>		<u>559,000</u>

Arithmetic

Average 5,6001,300

Sum of deviations squared ... 8,900,000

559,000

Arithmetic average of deviations squared

988,88962,111

Square root of arithmetic average of squares of deviations =

994.42

249.22

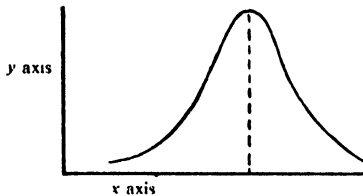
	Company A £	Company B £
Arithmetic average	5,600	1,300
Standard deviation	994.42	249.22
Coefficient of dispersion--		
S.D. × 100	994.42 × 100	249.22 × 100
Arithmetic Average	<u>5,600</u>	<u>1,300</u>
	= 17.75%	= 19.17%

The profits of these two companies are thus put on a *comparable* basis as far as variations from the average in each case are concerned.

SKEWNESS

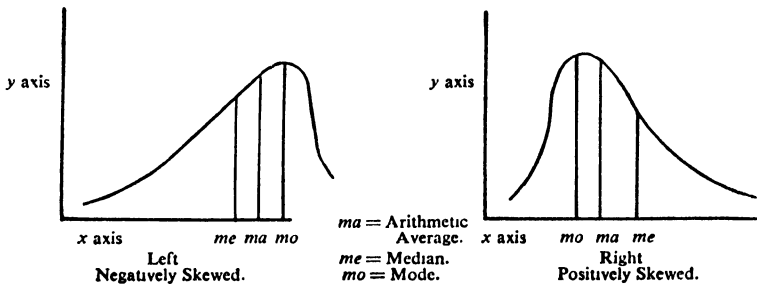
The various measures of dispersion outlined to date have been concerned with either the spread of the data or a consideration as to whether the individual items of the series vary from the average, whilst the coefficients of dispersion have illustrated how the *relative* variations may be expressed. None of these things, however, denotes skewness, i.e. how far and in what direction the distribution departs from a symmetrical distribution. The notion of skewness does not concern the accountant a great deal and but a passing reference is made here, simply to indicate that a measurement of this lack of symmetry exists.

Data which, when plotted in the form of a graph will result in a curve of the following type, is said to be a symmetrical distribution.



In such a curve as this the vertical line drawn through the centre of the curve denotes that the arithmetic average, the median and the mode, are all coincident, i.e. they have the same value, which is found at the point where the vertical line cuts the x axis.

Should, however, the curve be distorted, it may be skewed to the left or right as follows:



Common methods for measuring skewness are—

- (1)
$$\text{Skewness} = \frac{\text{Ma} - \text{Mo}}{\text{Standard Deviation.}}$$
- (2)
$$\text{Skewness} = \frac{3 (\text{Ma} - \text{Me})}{\text{Standard Deviation.}}$$
- (3)
$$\text{Skewness} = \frac{(\text{Q}_3 - \text{Me}) (\text{Me} - \text{Q}_1)}{\text{Quartile Deviation.}}$$

The above remarks on skewness are intended only to indicate the method whereby an index may be obtained so as to represent the skewness of a particular distribution and thus facilitate comparison with that of other distributions. Skewness is used so little in the analysis of accounting data that any further detail relating to it would be quite unwarranted in a book of this nature.

CHAPTER 4

CORRELATION

INTRODUCTION	57
MEASUREMENT OF CORRELATION	58
The Coefficient of Correlation.									
Illustrative Example 1.									
INTERPRETATION OF THE COEFFICIENT	61
Perfect Direct Correlation.									
Perfect Inverse Correlation.									
Complete Absence of Correlation.									
Varying Degrees of Correlation.									
Illustrative Example 2.									
SCATTER DIAGRAMS	69
Illustrative Example 3.									

INTRODUCTION

Quite often there is such a relationship between two series of figures that a change in one is accompanied by a change in a similar or opposite direction in the other. The change in the latter series may not be of the same degree as that in the former, but generally speaking it is often important for us to know whether between, say, series A and B there is a condition such that a movement in A results in a sympathetic movement in B, and if so, in what direction each movement is tending, i.e. whether in the same or opposite direction. These problems are concerned with the notion of Correlation.

Correlation deals with the observation and measurement of the relationship (if any) existing between statistical series. Where there is a movement in one series in a certain direction, and this is accompanied by a movement in the other series in the *same* direction, there is *direct* correlation between the two. If the movement in the former is accompanied by a movement in the latter in the reverse or opposite direction, then we say that as between the two series there is *inverse* correlation.

To illustrate, if an increase in the number of motor cars on the roads is accompanied by an increase in the number of accidents, then as between these two sets of conditions or objects there is direct correlation. On the other hand, should an increase in the number of policemen in the active force be accompanied by a

decrease in crimes committed, then as between these two sets there is inverse correlation. It should be noted, however, that while correlation appears to be present, the movements may not necessarily be related as to cause; e.g. an increase in road accidents may, on investigation, be found to be partly the result of a long period of bad weather, etc. The movements in both series may be the result of a third factor affecting each in some degree, rather than a sympathetic movement.

MEASUREMENT OF CORRELATION

In the interpretation of accounting data there are two main methods of testing two series for correlation:

- (a) by calculating the coefficient of correlation,
- (b) by preparing "scatter diagrams".

Correlation is of limited use in statistics as applied to accounting data and these two methods are generally considered to be those most applicable.

The Coefficient of Correlation

When considering the coefficient of correlation, two things are important, viz.:

- (a) the method by which it is calculated, and
- (b) the interpretation of the coefficient when once it is calculated.

In business statistics the most common method of calculating the coefficient of correlation is the *Pearsonian* method. This method, the formula for which is given later, involves the following steps in carrying it out:

1. Prepare a table as in Illustrative Example 1, setting up the original data in each series in columns 2 and 5.
2. Calculate the arithmetic average of each series.
3. Find the deviations of each of the items from the arithmetic average and insert in columns 3 and 6. These deviations are simply the result of the subtraction of each item from the arithmetic average of the series in which the item appears. It is important to see that the correct algebraic sign is affixed to the deviation.
4. Multiply each of the deviations in the first series (column 3) by the corresponding deviation in the second series (column 6) and insert the products in column 8, taking care to affix the proper algebraic sign resulting from the multiplication. It will be remembered that a negative quantity multiplied by a negative quantity results in a positive quantity and that a positive quantity multiplied by a negative quantity results in a negative quantity.

5. Square each of the deviations in columns 3 and 6 and insert the resultant figures in columns 4 and 7.
6. Add the columns containing the deviations squared and the products of deviations (columns 4, 7 and 8).
7. Find the standard deviation of each series by taking the square root of the total deviations squared divided by the number of items.
8. The coefficient of correlation is then given by an evaluation of the following identity:

$$\frac{\text{Total of Products of Deviations}}{\text{Number of items} \times \text{Stand. Dev. of Series A} \times \text{Stand. Dev. of Series B.}}$$

Illustrative Example 1

Calculate the coefficient of correlation between the following series:

Year ended 30th June	Sales of Implements £000's	Sales of Spare Parts £000's
1936	36	3.3
1937	38	3.4
1938	39	3.4
1939	42	3.5
1940	36	3.2
1941	34	3.0
1942	25	2.8
1943	22	2.9
1944	20	3.0
1945	28	2.5

Solution :

Year ended 30th June. (1)	Sales of Implements.			Sales of Spare Parts.			
	Original data. (2)	Deviations from Mean. (3) (x)	Deviations Squared. (4) (x ²)	Original data. (5)	Deviations from Mean. (6) (y)	Deviations Squared. (7) (y ²)	Product of Deviations. (8) (xy)
1936	36	4	16	3.3	0.2	0.04	+ 0.8
1937	38	6	36	3.4	0.3	0.09	+ 1.8
1938	39	7	49	3.4	0.3	0.09	+ 2.1
1939	42	10	100	3.5	0.4	0.16	+ 4.0
1940	36	4	16	3.2	0.1	0.01	+ 0.4
1941	34	2	4	3.0	—	0.01	— 0.2
1942	25	— 7	49	2.8	— 0.3	0.09	+ 2.1
1943	22	— 10	100	2.9	— 0.2	0.04	+ 2.0
1944	20	— 12	144	3.0	— 0.1	0.01	+ 1.2
1945	28	— 4	16	2.5	— 0.6	0.36	+ 2.4
<i>Total:</i>	320		530	31.0		0.90	+ 16.6
<i>Mean:</i>	32			3.1			

$$\text{Standard Deviation of Sale of Implements} = \sqrt{\frac{530}{10}} = \sqrt{53} = 7.28$$

$$\text{Standard Deviation of Sale of Spare Parts} = \sqrt{\frac{0.9}{10}} = \sqrt{0.09} = 0.30$$

$$\text{Coefficient of Correlation} = \frac{16.6}{10 \times 7.28 \times 0.30} = + 0.760$$

We will temporarily defer discussion of the interpretation of this particular coefficient since at this stage consideration is only being given to the method of computing it.

As indicated previously this coefficient is known as the Pearsonian Coefficient of Correlation and the method expressed in formula form is as follows:

$$r = \frac{\Sigma(xy)}{n\sigma_x\sigma_y}$$

where r = the coefficient of correlation.

x = the deviations of the items from the arithmetic average—in the first series.

y = the deviations of the items from the arithmetic average—in the second series.

n = the number of items, i.e. actually the number of pairs of items.

σ_x = the standard deviation of the first series.

σ_y = the standard deviation of the second series.

It is clear then that xy is the product of the deviations of both series and that $\Sigma(xy)$ represents the algebraic *sum* of all the products of the deviations of both series.

INTERPRETATION OF THE COEFFICIENT

Once the coefficient has been calculated the next and probably the most important matter is to interpret the meaning of the result obtained.

The following give some idea of the different types of series between which there is quite often a high degree of correlation:

- (a) Between the sales of machines and the sales of spare parts. In the case of a motor distributor for instance it can roughly be said that the greater his sales of cars, the greater will be his sales of spare parts. There is thus direct correlation between these two sets of things but the degree requires mathematical measurement along the above lines. It might be noted here that there is quite often the possibility of a time lag, so that, for instance, although increased sales of spare parts may follow increased sales of cars, the increase in the former may not be *immediately* apparent.
- (b) Between turnover and gross profit. Normally it is expected that the greater the value of sales, the greater profit will be made. The calculation of the coefficient of correlation of two series consisting of turnover and gross profit over a period of years should test whether in a particular case this is so or not.
- (c) Between the ages of workers and the times they take to perform a certain task. It may be that the older a worker, the longer he takes to complete the task. On the other hand, the results may disclose an absence of correlation.

These are only examples of the pairs of events between which it is often desired to ascertain whether or not correlation exists.

It is essential that in such cases the coefficient of correlation should make manifest two things, viz:

- (a) Whether the two sets of events are moving in the same *direction* or not, and
- (b) The *degree* of the relationship between them.

When properly interpreted the coefficient of correlation calculated according to the method described above should indicate both of these requirements.

The starting point in the interpretation of the coefficient is the fact that the limiting values of the coefficient are ± 1 . That is, the coefficient cannot exceed $+1$ in the positive (or *direct*) direction or -1 in the negative (or *inverse*) direction. If the coefficient is equal to $+1$ then as between the two series there is *perfect direct correlation*. If it is equal to -1 then as between the two series there is *perfect inverse correlation*. The algebraical sign indicates whether the two sets of events are moving in the same direction or not. If the sign is $+$, then they are moving in the same direction (i.e. the correlation is *direct*); if the sign is $-$, then they are moving in opposite directions (i.e. the correlation is *inverse*). The *degree* of the relationship is indicated by the *value* of the coefficient which in each particular case will lie somewhere between $+1$ and -1 .*

Before considering what the varying values between ± 1 might mean, it is proposed to illustrate how it is that perfect direct correlation results in a coefficient of $+1$ and perfect inverse correlation results in a coefficient of -1 .

Perfect Direct Correlation

For the purpose of illustrating how perfect direct correlation results in a coefficient of exactly 1, let us assume the following hypothetical series for sales and profits. In these series the profit figure in each case is 25 per cent. of sales so that as sales increase, so do profits. There is an increase in the same direction and also there is an increase in degree of the same value, since if sales are doubled then profit is doubled. It can be argued of course that this state of affairs could be ascertained merely by looking at the two series. This is so in this case which is particularly designed only for the purpose of demonstrating the theory underlying the formula. In many cases there may be a high degree of direct correlation (although it might not be perfect) but it could not be ascertained accurately unless the formula were applied. Take then the following series:

Sales	Profits
£	£
20	5
40	10
60	15
80	20
100	25
120	30
140	35

* If, however, the value of the coefficient is 0, this indicates a complete absence of correlation, see later.

The calculation of the coefficient of correlation is as follows:

Solution :

Sales.		Profits.	
Original Data.	Deviations from Mean. (x)	Original Data.	Deviations from Mean. (y)
20	60	5	15
40	40	10	10
60	20	15	5
80	—	20	—
100	20	25	5
120	40	30	10
140	60	35	15
Total:	560	140	700
Mean:	80	20	
	Deviations Squared. (x ²)	Deviations Squared. (y ²)	Products of Deviations. (xy)
	3,600	225	900
	1,600	100	400
	400	25	100
	—	—	—
	400	25	100
	1,600	100	400
	3,600	225	900
	11,200	700	2,800

$$\text{Standard Deviation of Sales} = \sigma_x = \sqrt{\frac{11,200}{7}}$$

$$= \sqrt{1,600}$$

$$= 40$$

$$\text{Standard Deviation of Profits} = \sigma_y = \sqrt{\frac{700}{7}}$$

$$= \sqrt{100}$$

$$= 10$$

$$\text{Coefficient of Correlation} = r = \frac{\sum(xy)}{n \sigma_x \sigma_y}$$

$$= \frac{2,800}{7 \times 40 \times 10}$$

$$= 1$$

Thus, in respect of two series between which we know there is perfect direct correlation, the coefficient or correlation is exactly + 1.

Perfect Inverse Correlation

To illustrate perfect inverse correlation assume the following series for costs and profits. From these it is clear that as costs increase profits decrease. The two series not only move in the opposite direction but they do so by the same degree, i.e. if costs are increased by £20, profits decrease uniformly by £1.

Costs	Profits
£	£
20	15
40	14
60	13
80	12
100	11
120	10
140	9

The calculation of the coefficient or correlation is as follows:

Costs.			Profits.			
Original Data.	Deviations from Mean. (x)	Deviations Squared. (x ²)	Original Data.	Deviations from Mean. (y)	Deviations Squared. (y ²)	Products of Deviations. (xy)
20	- 60	3,600	15	3	9	-180
40	- 40	1,600	14	2	4	- 80
60	- 20	400	13	1	1	- 20
80	-	-	12	-	-	-
100	20	400	11	1	1	20
120	40	1,600	10	2	4	80
140	60	3,600	9	3	9	180
Total:		11,200	84		28	-560
Mean:			12			

$$\text{Standard Deviation of Costs} = \sigma_x = \sqrt{\frac{11,200}{7}} = 40$$

$$\text{Standard Deviation of Profits} = \sigma_y = \sqrt{\frac{28}{7}} = 2$$

$$\text{Coefficient of Correlation} = r = \frac{\Sigma(xy)}{n \sigma_x \sigma_y} = \frac{-560}{7 \times 40 \times 2} = -1.$$

Thus in respect of two series between which we know there is perfect inverse correlation, the coefficient of correlation is exactly -1 .

Complete Absence of Correlation

If there is a complete absence of correlation between two series the coefficient of correlation will be 0.

Assume the following series:

Sales	Salaries
£	£
200	75
250	75
270	75
290	75
300	75
350	75

In this case as sales increase, there is no corresponding movement in salaries whatever. There is thus a complete *absence* of correlation. To demonstrate that the coefficient of correlation must be 0 is an easy matter for in the second series there is no deviation of each of the items from the average of the series. The "squares" of the deviations are, therefore, all 0, the products of the deviations of the two series are thus all 0, the numerator of the coefficient of correlation must be therefore 0, and so the coefficient itself is 0.

Varying Degrees of Correlation

Summarizing the above remarks on the interpretation of the coefficient of correlation, we have—

if $r = +1$, there is perfect direct correlation.

if $r = -1$, there is perfect inverse correlation.

if $r = 0$, there is a complete absence of correlation.

As indicated previously, however, perfect direct and inverse correlation will only be met with in rare cases. Usually if either exists, it is known beforehand and there is no need to compute the coefficient.

When two series are tested for correlation, the resultant coefficient will be somewhere between 1 and -1 . The following general indications afford a guide as to the interpretation of the coefficient of correlation.

Table of Interpretative Values

If the Coefficient lies between	It will generally indicate
0 and 0.30	Whether positive or negative, little evidence of direct or inverse correlation, and changes in items of the first series should not be taken to indicate what changes will result in the second series.
0.30 and 0.50	Whether positive or negative, there is some degree of direct or inverse correlation and changes in the items of one series should not necessarily be taken as a guide to the probable changes in the other.
0.50 and 0.75	Whether positive or negative, there is a decided correlation between the two series and changes in one series may be taken as a rough guide as to probable changes in the other series.
0.75 and 0.90	Whether positive or negative, there is a fairly high degree of direct or inverse correlation between the two series. Estimations of items of one series may be made from known values of the other series with reasonable accuracy.
0.90 and 1.00	Whether positive or negative, there is a high degree of direct or inverse correlation. Estimations of items of one series may be made from known values of the other series with substantial accuracy particularly where the coefficient is greater than 0.95.

Another illustrative example will now be worked to test the interpretative value to be placed on the coefficient.

Illustrative Example 2

Calculate the coefficient of correlation between the following series. What does this coefficient indicate?

Year	Series A	Series B
1	8,000	300
2	7,500	350
3	9,000	400
4	9,600	450
5	10,400	500

Solution:

	Series A.			Series B.			
	Original Data.	Deviations from Mean. (x)	Deviations Squared. (x ²)	Original Data.	Deviations from Mean. (y)	Deviations Squared. (y ²)	Product of Deviations. (xy)
1	8,000	- 900	810,000	300	- 100	10,000	90,000
2	7,500	-1,400	1,960,000	350	- 50	2,500	70,000
3	9,000	100	10,000	400	0	—	—
4	9,600	700	490,000	450	50	2,500	35,000
5	10,400	1,500	2,250,000	500	100	10,000	150,000
<i>Total:</i>	44,500		5,520,000	2,000		25,000	345,000
<i>Mean:</i>	8,900			400			

$$\text{Standard deviation of A} = \sqrt{\frac{5520000}{5}} = \sqrt{1104000} = 1050.71$$

$$\text{Standard deviation of B} = \sqrt{\frac{25000}{5}} = \sqrt{5000} = 70.71$$

$$\text{Coefficient of Correlation} = \frac{345000}{5 \times 1050.71 \times 70.71} = 0.928$$

The coefficient of correlation being greater than 0.9, there is thus a high degree of direct correlation between the two series and an estimation of items of one series may be made from known values of the other series, with substantial accuracy.

SCATTER DIAGRAMS

A test for correlation between two series may be made by means of a "scatter diagram". In some cases the preparation of a graph or diagram may be more laborious than the mathematical calculation of the coefficient and it is not always used. Furthermore, it does not give the mathematical precision of the former method. The great value of a scatter diagram, however, is the pictorial presentation which often is more easily understood by those for whom it is prepared.

Illustrative Example 3

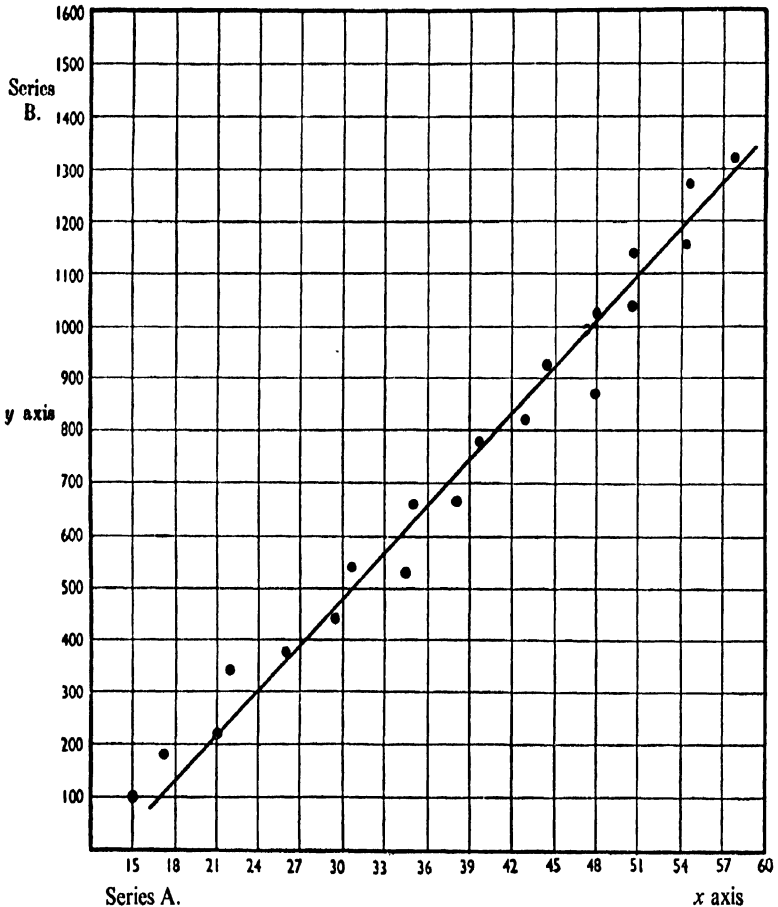
Prepare a scatter diagram to illustrate the suspected correlation between the following series:

Series A	Series B
15	100
17	185
21	230
22	350
26	385
29	440
31	540
34	530
35	670
38	670
40	785
43	815
44	920
48	875
48	1,010
50	1,030
50	1,150
55	1,160
55	1,275
58	1,310

To construct this diagram the first series is placed on the x axis and the second series on the y axis and the various points plotted according to a convenient scale.

It will be noticed that the dots tend to follow a line sloping upwards from left to right. This indicates a high degree of direct correlation. If the correlation were highly inverse, the dots would tend to follow a line sloping downwards from left to right. If no correlation existed, the dots would be spread over the graph and would not follow any particular pattern.

Actually the dots in the diagram follow more of a path than a line and the closer the two series are to perfect correlation,



SCATTER DIAGRAM
Showing High Degree of Direct Correlation.

the narrower becomes the path until, if perfect correlation does exist, the path of dots converges into the straight line.

The straight line drawn through the dots has been drawn free-hand, i.e. it is simply a line such that it describes what appears to be the direction of movement. The insertion of this "*line of best fit*" enables estimations to be made of items in one series from known values of items in the other series.

For instance, if it is required to find what value in Series B would correspond to the value 42 in Series A, all that is necessary is to trace from point 42 on the x axis to the straight line and from there across to the y axis and read off the value, 830.

CHAPTER 5

TREND AND SEASONAL VARIATIONS

INTRODUCTION	71
Necessity for analysis of time series.									
CLASSIFICATION OF CHANGES IN TIME SERIES							73
Secular Trend.									
Seasonal Variations.									
Cyclical Movements.									
Irregular Variations.									
THE MEASUREMENT OF SECULAR TREND						77
1. Freehand Curve Fitting Method.									
2. Semi-averages Method.									
Illustrative Example 1.									
3. Moving Averages Method.									
Illustrative Example 2.									
4. Least Squares Method.									
Illustrative Example 3.									
Illustrative Example 4.									
Elimination of the Secular Trend.									
THE MEASUREMENT OF SEASONAL VARIATIONS							90
1. Monthly Averages Method.									
Illustrative Example 5.									
2. Moving Averages Method.									
Removal of Seasonal Variations.									
Illustrative Example 6.									

INTRODUCTION

In this chapter we are concerned with—

- (a) the description, and
- (b) the measurement

of movements and changes which occur or appear in a statistical series over a period of time. Such a series is known as a time series. In Chapter 6 when considering the general classification of statistical series it was stated that a time series was one in which the data were grouped according to time of occurrence. A simple example of a time series is as follows:

Telephones (N.S.W.)

30th June	Telephone Lines Connected	Telephone Instruments Connected
1940	197,046	268,216
1941	205,097	280,161
1942	206,103	285,243
1943	210,112	295,266
1944	217,345	308,016
1945	225,832	320,198

Source: N.S.W. Statesman's Year Book for 1947.

In the above example, the independent variable—time—is constituted by a series of consecutive years and the dependent variables, which fluctuate in time, are constituted by the number of telephone lines and telephone instruments connected at 30th June in each year. Similar tables might be prepared for, say, monthly sales, weekly price quotations, daily or hourly returns of electricity consumption, etc. In each case, the various values of the dependent variables—sales, prices, electricity consumption—would be related to definite points in time.

Necessity for Analysis of Time Series

A simple illustration will show the necessity for investigating the general tendency of a series to move in a certain direction, i.e. for investigating the trend of a series.

Assume that the following data represents the net profits earned by two companies over a period of ten years.

Year	Net Profit	
	Company A	Company B
	£	£
1941	5,000	28,650
1942	5,200	26,420
1943	6,700	23,290
1944	6,850	21,985
1945	8,405	19,200
1946	7,200	18,400
1947	8,750	18,300
1948	9,270	20,870
1949	10,050	17,250
1950	11,090	15,490

These series exhibit two distinct and opposite movements. In Company A, profits have steadily increased over the 10 years, and there is an *upward trend*, or a growth. In Company B, profits have steadily decreased, showing that there is a *downward trend*, or a decline.

These figures for a period of years supply information which would not be forthcoming from a consideration of the profit for a particular year, or, further, from a comparison of the profit for a particular year with that of the previous year. In Company A,

for instance, the profit for 1946 was lower than that for 1945 despite the fact that over the period of ten years profits have shown a definite trend of growth. If in 1946 a comparison had been made with the profit for 1945, the decrease noted, and the assumption made that profits were decreasing, such assumption would not have been correct. Again, in the case of Company B, there is a downward trend of profits despite the increase in 1948 as compared with 1947.

The study of trend is important, not only for investigating the progress of a particular enterprise but also for reviewing business conditions generally. Quite often a particular enterprise will prosper only if it "feels the pulse" of the trade and correctly anticipates changes and movements within it as indicated by trends. Whilst trend is the study of past events or performances occurring throughout the period of time taken, its purposes are to determine the *direction* in which the data are moving and assist business forecasting. Most business executives attempt to forecast future demands of the consumer, future possibilities of supplies, future rises or falls in prices, and so on. Thus, an enterprise may stock up heavily if it considers that costs of materials will rise or supplies will become scarce and, on the other hand, it may attempt to rid itself of its stocks as quickly as possible if the prices at which it will be able to sell are likely to fall drastically. These are only instances of the circumstances in which businessmen require a fairly sound idea as to how events are moving, i.e. as to the direction in which the trend of business generally is moving.

CLASSIFICATION OF CHANGES IN TIME SERIES

Generally speaking there are four classes of changes or movements in a time series. These are as follow:

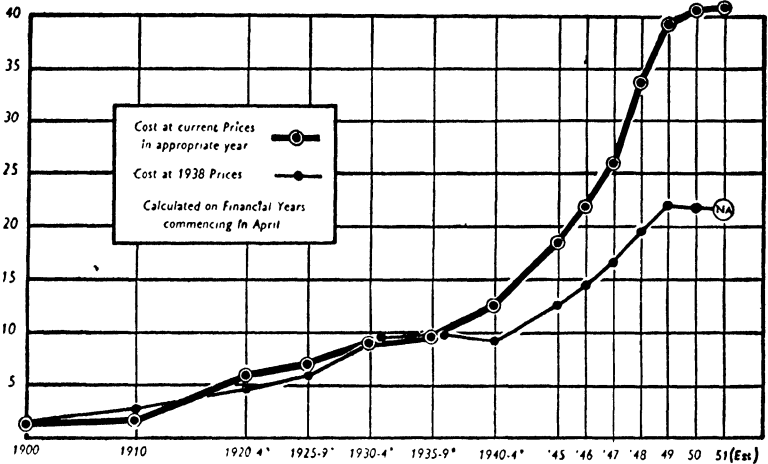
1. Secular or Long-term Trend.
2. Seasonal Variations.
3. Cyclical Movements.
4. Irregular Variations.

1. Secular Trend

The secular trend is a long-term trend. That is, it is the growth or decline exhibited by the series over a period of years. When discussing the movement of profits of companies A and B on page 72, what we were really considering was the secular trend of the data, i.e. its tendency to move in a certain direction over the period of ten years. Generally, secular trend should not be looked for in data extending over a period of less than ten years, unless there is special reason for doing so. In the following diagram extracted from the *Barclay's Bank Review* it will be seen that the

COST OF SOCIAL SERVICES IN GREAT BRITAIN

£ PER HEAD OF POPULATION



trend of social service costs per head of population has been steadily upwards both in terms of money and of purchasing power in the fifty years under review.

2. Seasonal Variations

Seasonal variations are fairly regular movements occurring within a period of twelve months. As the name implies the variations are due to the differences in seasons and social festive occasions. Thus, ice cream sales are greater in summer than in winter, sales of winter clothing are greater in winter than in summer, businesses in holiday resorts will do the great bulk of their trade in the holiday seasons and little else during the rest of the year, and so on. Frequently, however, the term seasonal variation is given a much wider meaning, e.g. there may be certain days of the week when sales are greater than the other days, and there are fluctuations in electricity consumption throughout each period of 24 hours, etc.

Social festive occasions, such as Christmas, Easter, Mother's Day, etc., must be considered also in the trend of seasonal variations. (See *post* under Methods of Measuring Seasonal Variation.)

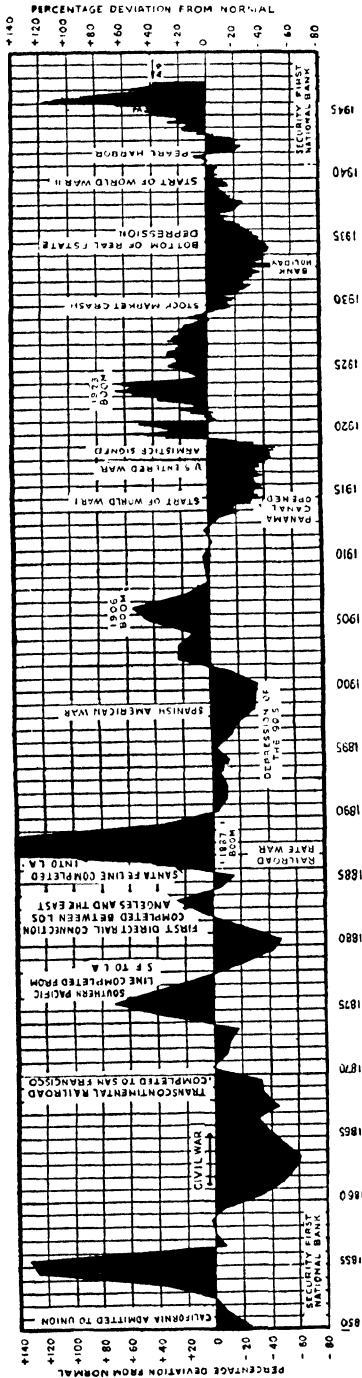
3. Cyclical Movements

It is now generally recognized that business conditions tend to move in cycles. These cycles are characterized by the tendency of business to recover from a depression and climb to prosperity, decline again and descend to another period of depression and so on. Business cycles are irregular and they vary in duration and intensity.

The diagram on page 76 is a graphical representation of a Real Estate Cycle reprinted here by permission of the Security-First National Bank of Los Angeles. It will be observed from the graph that there have been six points of highest level or peaks, viz., those occurring in 1855, 1875, 1887, 1906, 1923 and 1946. Between each of these peaks are low points at 1863, 1880, 1900, 1918 and 1933. Thus real estate sales activity in Los Angeles County between 1850 and 1946 exhibits a fairly regular cyclical pattern. Other forms of activity do not always exhibit the regularity manifested in this particular case. However, the Monthly Summary of Business Conditions from which the chart is extracted, points out that, "in considering the possible course of real estate activity in the period ahead, readers are cautioned against making detailed forecasts on the basis of historical precedent. Many things have affected the specific character and timing of individual real estate cycles in the past, both in this and other areas; among these are (1) wars, (2) major catastrophes, such as earthquakes, fires and floods, (3) discovery of important oil fields or other natural resources, (4) completion of railroads or harbour facilities opening up an area to extensive development, (5) major waves of population influx, (6) financial crises, and cyclical swings in business activity."

These general considerations apply with the same force to the British economy (see further under Irregular Variations).

Trend and Seasonal Variations



REAL ESTATE CYCLE IN LOS ANGELES COUNTY
 PERCENTAGE DEVIATION FROM "NORMAL."

4. Irregular Variations

The effect of possible irregular or unusual disturbances on trends and trade cycles cannot be ignored. It was seen, when considering cyclical movements, that such random variations exerted a major influence on the timing and extent of fluctuations, such an influence in fact, that since they are to be expected in the future, specific predictions as to real estate activity (the case under consideration) would be meaningless.

Irregular variations may be caused by wars, floods, droughts, earthquakes, fires, mineral discoveries, strikes, large-scale immigration or emigration, institution of important public utilities, and so on. Some of these events have exerted great influence on economic conditions, particularly mineral discoveries, droughts, strikes and wars. Certain forms of activity are more prone than others to reflect within themselves the effects of extraneous events.

THE MEASUREMENT OF SECULAR TREND

The measurement of trend involves the determination of a trend line such that it is representative of the normal values which would have been recorded had no forces other than those of normal growth or decline been operating. Thus the problem involves the elimination from the data of cyclical movements, seasonal variations and irregular variations. A *trend* value is thus a *normal* value. The problem of the measurement of seasonal variations is dealt with separately later in this chapter.

To fit a trend line to data it is wise to plot the data on a graph first. The resultant chart may indicate such a simple and straightforward movement that the trend line may be fitted quite easily by freehand (see later) and there will not be any need to compute trend values by any other mathematical method. The value of the preparation of a graph as a preliminary consideration to fitting a trend is also apparent where there are any extreme items in the series. Such extreme items will show up readily in a plotted chart and their effect can be guarded against. The selection of the period of years over which the data extends is important. As indicated previously such period should not be less than ten years and further it should not start or end with an unusual period of activity such as a boom or a depression.

The most common methods of fitting a line of trend to data are:

1. Freehand Curve-fitting Method.
2. Semi-Averages Method.
3. Moving Averages Method.
4. Least Squares Method or Method of Mathematical Curve Fitting.

An outline and brief explanation of each of these methods will now be considered.

1. Freehand Curve-fitting Method

This method of fitting a trend line to data simply involves the drawing by freehand of a line through the data in such a way that the line describes what appears to be the general direction of movement of the data. The drawing of the line need not be strictly freehand and a ruler, etc., may be used if necessary, but the method does not involve any mathematical calculation whatever. If the line is drawn with care and good judgment is exercised in tracing it and in choosing its direction, it may closely approximate to a trend line fitted by mathematical calculation.

In the diagrams on page 79 a straight line has been superimposed on the right hand diagram, and a curved line on the left hand diagram, over the lines representing Britain's import trade with certain areas.

These trend lines* illustrate or describe the general direction in which imports have moved during the period 1948/51. As to whether this trend is an accurate description of the movement depends on mathematical calculations, which may modify the line in some respects. It should be noted, however, that the trend line eliminates or ignores the bulges occurring in the years 1949 and 1950.

2. The Semi-Averages Method

Under this method of fitting a trend line, the data is divided into two equal parts and each half is averaged. The averages so obtained are then plotted on the graph at the points which represent the centre of each of the halves of the original data. A straight line is then drawn through the two points so plotted and this represents the line of trend.

An illustration will make the procedure clear.

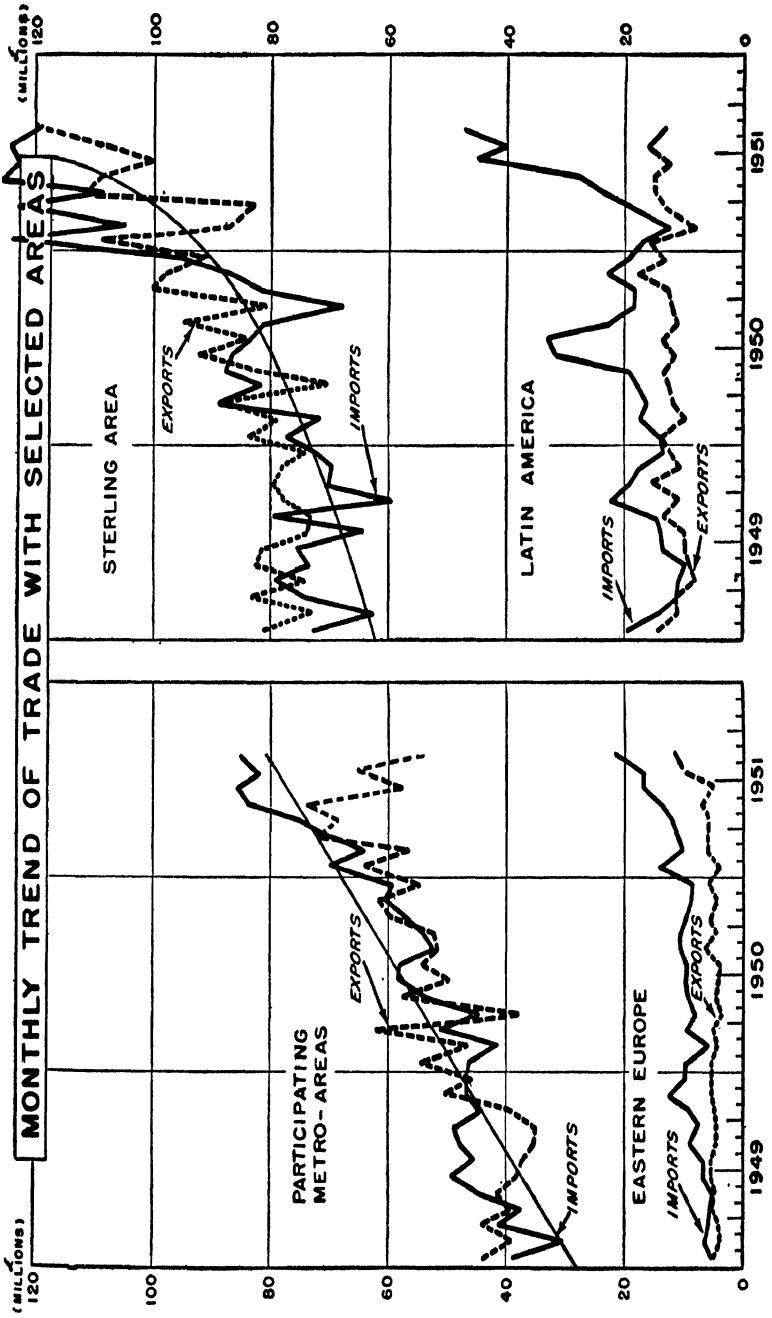
Illustrative Example 1

The sales of the Alpha Co. Ltd. for the period 1939–1948 are as follow:

Year	Sales £000s
1939	18
1940	22
1941	45
1942	50
1943	60
1944	62
1945	60
1946	63
1947	70
1948	90

Plot the graph of the above series and fit the line of trend by the Semi-Averages Method.

*These trend lines have been inserted by the author and were not originally part of the chart as extracted from the Statistical Bulletin of the Economic Co-operation Administration.



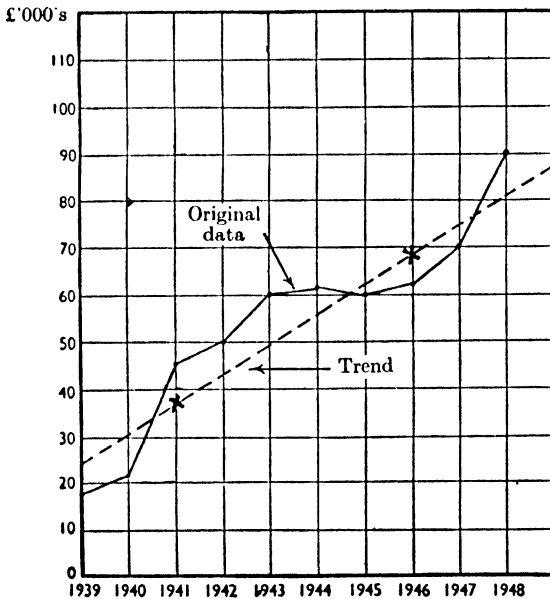
Solution:

Year	Sales	Totals	Arithmetic Averages
1939	18	195	39
1940	22		
1941	45		
1942	50		
1943	60		
1944	62	345	69
1945	60		
1946	63		
1947	70		
1948	90		

The data is divided into two equal parts for the periods 1939–1943 and 1944–1948. Each half is totalled and the amount inserted in the totals column. The arithmetic average of each of these totals is next obtained by dividing each total by the number of years it represents, i.e. in this case five each. Note that the totals and their averages are placed opposite the centre year in each half (i.e. opposite 1941 and 1946 respectively).

The next step is to plot the original data and join the points to form a graph of the sales over the period. Then the value 39 is plotted against the year 1941 and the value 69 against the year 1946 to give the two crosses marked on the graph. A straight line is then drawn through these two points and this is the line of trend.

The graph with its line of trend would appear as follows:



SALES OF ALPHA CO. LTD., 1939–1948.
Trend ascertained by Semi-Averages Method.

3. Moving Averages Method

By this method of fitting a trend line to the data, the trend ordinates (which are plotted so as to determine the line of trend), are obtained by successively averaging the data in cycles of 3, or 5, or 7, or 9, etc., years. A cycle is really a number of years to be averaged and the selection of a suitable cycle is the first requirement when using this method. For instance, if a cycle of three years were selected then the data would be successively averaged in groups of three.

As indicated previously, a trend line should eliminate cyclical and irregular variations. Whether a three-year, or five-year, or seven-year, or nine-year cycle is chosen depends upon how long a period is necessary to remove the effects of these cyclical and irregular variations.

The following example illustrates the method of computing the moving averages for purposes of plotting the line of trend.

Illustrative Example 2

The value of production of the Gamma Factory has varied throughout the period indicated as follows:

Year	Production £000s
1935	16
1936	17
1937	21
1938	31
1939	35
1940	39
1941	43
1942	53
1943	57
1944	61
1945	71

You are required to:

- (i) Calculate the trend values for the above series, using a three-year moving average.
- (ii) Graph the above data.
- (iii) Graph the trend.

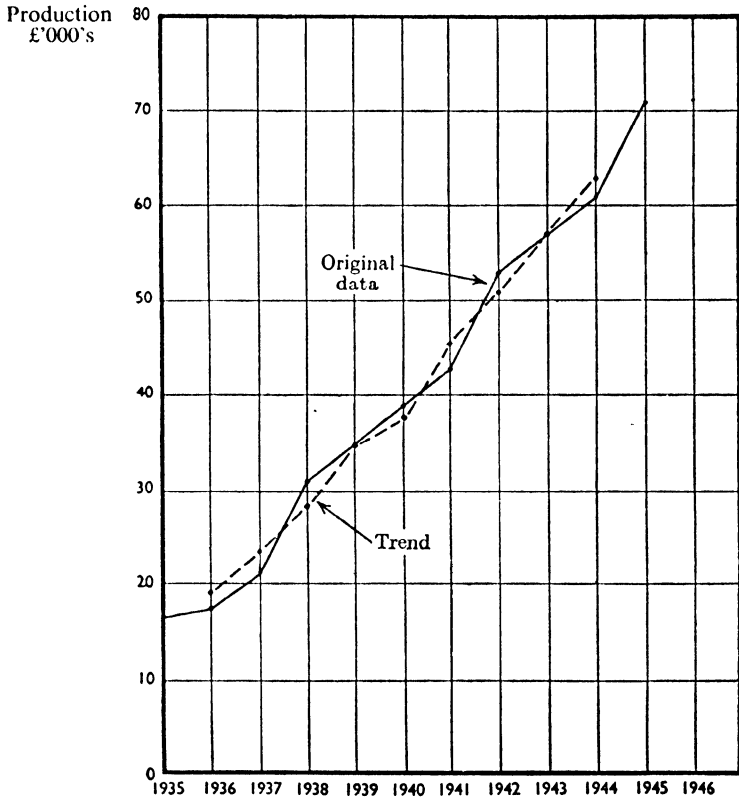
Solution:

Year	Production £000s	3-year Moving Totals	3 years Moving Average
1935	16		
1936	17	54	18
1937	21	69	23
1938	31	87	29
1939	35	105	35
1940	39	117	39
1941	43	135	45
1942	53	153	51
1943	57	171	57
1944	61	189	63
1945	71		

The figures in the "3 years moving totals" column were obtained by (a) adding the first three years' items together, (b) dropping the first item and adding the next three items, (c) dropping the first and second items and adding the next three, and so on. In other words, the first total is obtained by adding the values for 1935, 1936 and 1937; the second total obtained by adding the values for 1936, 1937 and 1938; the third total obtained by adding the values for 1937, 1938 and 1939, and so on. The moving totals and the moving averages (these latter being obtained by dividing the moving total by the cycle period, which is, in this case, 3) are placed against the middle year of the 3 years averaged.

The moving averages are then plotted on the graph against the years opposite which they appear and when these points are joined the resultant line is the line of trend.

In the following diagram the solid line is the graph of the original data and the dotted line is the trend line.



PRODUCTION—GAMMA FACTORY, 1935-1945

Trend ascertained by three-year Moving Average.

If the trend line is to be based on a five-year moving average, then each successive five years' values are averaged and computed on the same principles as above. The same applies to any other cycle used.

It should be noted that in this particular method, the trend line cannot be brought up to date since the averages are representative of the middle point of each cycle.

Further, if the cycle taken is an even number of years then the average will lie between the *two* middle years of the cycle and thus the averages will require to be recentred so as to bring them opposite a particular middle year. For instance, assume that in the previous example it was required to find the line of trend basing it on a *four*-year moving average. The averages would be computed as follows:

(1)	(2)	(3)	(4)	(5)	(6)
Year	Production	4-year Moving Total	4-year Moving Average	2-year Moving Totals of Column 4 added in pairs	4-year Moving Average Centred
1935	16				
1936	17				
1937	21	85	21.25	47.25	23.63
1938	31	104	26.00	57.50	28.75
1939	35	126	31.50	68.50	34.25
1940	39	148	37.00	79.50	39.75
1941	43	170	42.50	90.50	45.25
1942	53	192	48.00	101.50	50.75
1943	57	214	53.50	114.00	57.00
1944	61	242	60.50		
1945	71				

It will be seen from the above that the items in Column (6) are now opposite a particular year and thus can be plotted to give the line of trend.

4. The Least Squares Method

The Least Squares Method of finding the trend ordinates (sometimes referred to as the method of Mathematical Curve Fitting) is slightly more involved than the previous methods and requires some mathematical calculation. If it is to be assumed that the

line of trend is to be a *straight line* then such trend will have a formula:

$$y = mx + b$$

because, mathematically, any straight line is reducible to such a formula. In this formula y and x are variables and further y is termed the dependent variable because its value depends upon what value is given to x . The symbol x itself is called the independent variable. It will be noticed that the assumption that the line of trend is a straight line presumes that the series is in arithmetic progression in the line of trend. For a short explanation of arithmetic progression see Chapter 6. Now, the problem involved is to find the *straight line* which best describes the data. The method of least squares, or it is really more correct to say, the principle of least squares, aids in determining that line. Once the trend values are found they may be graphed as in the previous methods, to give the straight line forming the trend of the data.

From the formula—

$$y = mx + b$$

it will be seen that there are two constants, viz., m and b .

In order to determine what these constants are it is necessary to solve two equations. These are of the form:

$$\begin{aligned} \Sigma(y) &= nb & + m\Sigma(x) \\ \Sigma(xy) &= b\Sigma(x) & + m\Sigma(x^2). \end{aligned}$$

These are two normal equations arising from the above formula and it is necessary now to substitute values in these equations for:

- $\Sigma(y)$, which is the sum of the values of y .
- $\Sigma(x)$, which is the sum of the values of x .
- $\Sigma(xy)$, which is the sum of the products of each of the x 's and its respective y .
- $\Sigma(x^2)$, which is the sum of the squares of the value of x .
- n , which is the number of years or items as the case may be.

To see how we get the value for each of these symbols we will work through an actual example.

Illustrative Example 3

It is required to compute the trend ordinates for the following data so that the trend will conform to the equation of $y = mx + b$.

Year	Sales £
1944	1,200
1945	1,300
1946	1,250
1947	1,400
1948	1,450
1949	1,500
1950	1,600

Solution:

It is first necessary to prepare a table with six columns ruled as follows:

(1) Year	(2) x	(3) Sales y	(4) xy	(5) x^2	(6) Trend Ordinates
1944	0	1,200	0	0	1,194
1945	1	1,300	1,300	1	1,258
1946	2	1,250	2,500	4	1,322
1947	3	1,400	4,200	9	1,386
1948	4	1,450	5,800	16	1,450
1949	5	1,500	7,500	25	1,514
1950	6	1,600	9,600	36	1,578
Totals	$\Sigma(x) = 21$	$\Sigma(y) = 9,700$	$\Sigma(xy) = 30,900$	$\Sigma(x^2) = 91$	

The figures in column (6), i.e. the trend ordinates, represent the final answer to the problem and of course are inserted last. In column (2) the years are consecutively numbered, the first year starting from 0 being the origin year. This column represents the independent variable x . The original data is inserted in column (3) and becomes the dependent variable (y) of the equation. In column (4) are inserted the products of x and y , i.e. the products of the two variables. Column (5) carries the squares of the independent variable x .

The respective columns are then totalled as indicated.

It is now clear how we get the values of the symbols for insertion in the two normal equations. They are:

$$\begin{array}{r}
 \Sigma(y) \qquad \qquad 9,700 \\
 \Sigma(x) \qquad \qquad \qquad 21 \\
 \Sigma(xy) \qquad \qquad 30,900 \\
 \Sigma(x^2) \qquad \qquad \qquad 91 \\
 n \qquad \qquad \qquad \qquad \qquad 7
 \end{array}$$

Substituting these values in the two normal equations—

$$\begin{array}{l}
 \Sigma(y) = nb \quad + \quad m\Sigma(x) \\
 \Sigma(xy) = b\Sigma(x) \quad + \quad m\Sigma(x^2)
 \end{array}$$

We have:

$$\begin{array}{l}
 9,700 = 7b \quad + \quad 21m \dots\dots\dots(1) \\
 30,900 = 21b \quad + \quad 91m \dots\dots\dots(2)
 \end{array}$$

Thus, it is necessary for us to solve these simultaneous equations to find the values of b and m .

Multiplying equation (1) by 3 and leaving equation (2) as it is we have:

$$\begin{array}{l}
 29,100 = 21b \quad + \quad 63m \dots\dots\dots(3) \\
 30,900 = 21b \quad + \quad 91m \dots\dots\dots(4)
 \end{array}$$

Subtracting equation (3) from equation (4):

$$\begin{aligned} 1,800 &= 28m \\ \therefore m &= \frac{1,800}{28} = 64 \text{ (approx.)} \end{aligned}$$

To find the value of b substitute 64 for m in equation (1) and we have:

$$\begin{aligned} 9,700 &= 7b + (21 \times 64) \\ &= 7b + 1,344 \\ \therefore 7b &= 9,700 - 1,344 \\ \therefore b &= \frac{8,356}{7} = 1,194 \text{ (approx.)} \end{aligned}$$

The values of m and b are therefore:

$$\begin{aligned} m &= 64 \\ b &= 1,194 \end{aligned}$$

and substituting these in the equation—

$$y = mx + b$$

we have

$$y = 64x + 1,194.$$

To ascertain the trend ordinates is now quite a simple matter. All that is necessary is to give x each of the seven different values it has in column (2) and we get seven different values for y which are the trend ordinates.

$$\begin{array}{ll} \text{Where } x = 0 & \text{then } y = (64 \times 0) + 1,194 = 1,194 \\ \text{,, } x = 1 & \text{,, } y = (64 \times 1) + 1,194 = 1,258 \\ \text{,, } x = 2 & \text{,, } y = (64 \times 2) + 1,194 = 1,322 \\ \text{,, } x = 3 & \text{,, } y = (64 \times 3) + 1,194 = 1,386 \\ \text{,, } x = 4 & \text{,, } y = (64 \times 4) + 1,194 = 1,450 \\ \text{,, } x = 5 & \text{,, } y = (64 \times 5) + 1,194 = 1,514 \\ \text{,, } x = 6 & \text{,, } y = (64 \times 6) + 1,194 = 1,578 \end{array}$$

These trend ordinates are then inserted in column (6) in the table against their respective years. When they are plotted on the graph and joined together the result will give the line of trend.

Alternative Method

Sometimes the least squares method is simplified by assuming that x lies at the middle of the period; Σx thus becomes zero. The normal equations then take the form—

$$\begin{aligned} \Sigma(y) &= nb \\ \Sigma(xy) &= m\Sigma(x^2) \end{aligned}$$

From the first equation it can be seen that

$$b = \frac{\Sigma y}{n} \text{ (} b \text{ is thus the arithmetic average of the series)}$$

and from the second equation—

$$m = \frac{\Sigma(xy)}{\Sigma(x^2)}$$

If this form were adopted, the working of Illustrative Example 3 would then appear as follows:

Year	x	y	xy	x ²	Trend Ordinates
1944	-3	1,200	-3,600	9	1,194
1945	-2	1,300	-2,600	4	1,258
1946	-1	1,250	-1,250	1	1,322
1947	0	1,400	—	—	1,386
1948	1	1,450	1,450	1	1,450
1949	2	1,500	3,000	4	1,514
1950	3	1,600	4,800	9	1,578
		Σ(y) = 9,700	Σ(xy) = 1,800	Σ(x ²) = 28	

$$b = \frac{\Sigma(y)}{n} = \frac{9,700}{7} = 1,386.$$

$$m = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{1,800}{28} = 64.$$

Substituting these values in the formula, $y = mx + b$ for say the sixth year, we have:

$$\begin{aligned} y &= (64 \times 2) + 1,386 \\ &= 128 + 1,386 \\ &= 1,514 \text{ (i.e. the trend ordinate for year six).} \end{aligned}$$

The trend ordinates for the other years would be calculated on the same principle.

In illustrating this method of least squares, the equation:

$$y = mx + b$$

and the normal equations derivable from it have been assumed. It is not within the scope of this book to explain the mathematical theory underlying these equations but the method is probably the most accurate one of inserting a line of trend in graphed data, *provided* that a *straight line* will serve as the most satisfactory trend. Further, an advantage of this method is that an extension of the trend line for say one or two years ahead will generally give fairly reliable estimates for these years. For instance, in the example just considered, if we read off on the y axis the value of the point where the extended line of trend cuts the vertical ordinate through the year 1951, such value would be an estimated figure for sales for 1951. Care must be exercised however in relying on such estimates and the comments made earlier under irregular variations should be noted.

In addition the following factors must be taken into consideration before reliance is placed upon a forecast as indicated by the line of trend:

- (a) If the data used were obtained by sampling, then it only represents a portion of the whole data, but since we wish to use the sample to draw conclusions about the data as a whole the limitations of sampling should be understood. (See Chapter 7.)

- (b) The data and the statistical analysis to which it has been subjected are all quantitative measurements. Social and economic forces may modify any interpretation of the results obtained by statistical analysis.
- (c) Mathematical methods are rigid. They tend to treat data on the assumption that they will conform to mathematical principles.
- (d) Statistical methods are not always adequate in that they are in some cases far too refined to interpret properly the data they handle. It is for this reason that a line of trend should not be fitted to data extending over a small number of years.

In the following example, the trend ordinates are calculated in a series dealing with an even number of years.

Illustrative Example 4

Year	x	Sales y	xy	x^2	Trend Ordinates
1946	0	1,200	—	0	1,200
1947	1	1,300	1,300	1	1,260
1948	2	1,250	2,500	4	1,320
1949	3	1,400	4,200	9	1,380
1950	4	1,450	5,800	16	1,440
1951	5	1,500	7,500	25	1,500
Totals	$\Sigma(x) = 15$	$\Sigma(y) = 8,100$	$\Sigma(xy) = 21,300$	$\Sigma(x^2) = 55$	

The two normal equations are:

$$\begin{aligned}\Sigma(y) &= nb + m\Sigma(x) \\ \Sigma(xy) &= b\Sigma x + m\Sigma(x^2)\end{aligned}$$

Substituting the above values in these equations we have:

$$\begin{aligned}8,100 &= 6b + 15m \\ 21,300 &= 15b + 55m\end{aligned}$$

Solving these:

$$\begin{aligned}m &= 60 \\ b &= 1,200\end{aligned}$$

Substituting these values in $y = mx + b$, for the first year we have:

$$\begin{aligned}y &= (60 \times 0) + 1,200 \\ &= 1,200\end{aligned}$$

The trend ordinates for the other years would be calculated on the same principle.

By the *alternative method*, taking the centre year as being notionally between 1948 and 1949, thus meaning that these two years are

each 0.5 of a year from the "centre year", the solution would be as follows:

Year	x	Sales y	xy	x ²	Trend Ordinates
1946	-2.5	1,200	-3,000	6.25	1,200
1947	-1.5	1,300	-1,950	2.25	1,260
1948	-0.5	1,250	-625	0.25	1,320
1949	+0.5	1,400	700	0.25	1,380
1950	+1.5	1,450	2,175	2.25	1,440
1951	+2.5	1,500	3,750	6.25	1,500
Totals	$\Sigma(x) = 0$	$\Sigma(y) = 8,100$	$\Sigma(xy) = 1,050$	$\Sigma(x^2) = 17.5$	

$$b = \frac{\Sigma(y)}{n} = \frac{8,100}{6} = 1,350$$

$$m = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{1,050}{17.5} = 60$$

Substituting these values in the formula $y = mx + b$, for the first year, we have:

$$y = (60 \times -2.5) + 1,350 = 1,200.$$

The trend ordinates for the other years would be calculated on the same principle.

Elimination of the Secular Trend

In order to eliminate the influence which the trend has on the data, the method used is to express the data as a percentage of the trend ordinate which corresponds with the items of original data.

The following table reproduces the information contained in columns (1), (3) and (6) of the table in Illustrative Example 3 and in addition shows the computations for the percentages of each of the original items to its trend ordinate (column (4) in the table below). Also, if from each of these percentages 100 is subtracted, then the result will be the percentage deviation from the line of trend (column (5)).

(1) Year	(2) Sales	(3) Trend Ordinates	(4) Percentage of Col. (2) to Col. (3)	(5) Deviations from trend %
1944	1,200	1,194	100.50	+0.50
1945	1,300	1,258	103.33	+3.33
1946	1,250	1,322	94.55	-5.45
1947	1,400	1,386	101.01	+1.01
1948	1,450	1,450	100.00	0
1949	1,500	1,514	99.07	-0.93
1950	1,600	1,578	101.39	+1.39

The interpretation of the figures in column (5) is that in 1946, for example, the actual value of sales is 5.45 per cent. below what would have been recorded if the series had been affected only by forces of normal growth.

MEASUREMENT OF SEASONAL VARIATIONS

As indicated earlier in the chapter, seasonal variations are fairly regular movements occurring generally within a twelve months' period. It can generally be said that if a time series exhibits seasonal variations then each month has a typical value in relation to the values for the other months of the year. The problem with which we are concerned here is the procedure by which the typical value for each month is ascertained.

There are several methods used for measuring seasonal variations, but the two common ones which will be considered here, are:

- (a) the monthly average method,
- (b) the moving average method.

Monthly Average Method

The purpose of this method is to calculate a seasonal index for each month so that we can tell how far above or below the monthly average for the year the individual monthly averages are.

Illustrative Example 5

Using the "monthly average" method, calculate a seasonal index of Sales from the following data:

Month	Sales in £'s						
	1945	1946	1947	1948	1949	1950	1951
January ...	200	185	190	175	170	195	215
February ...	180	170	175	160	150	170	185
March... ..	130	125	140	125	115	140	135
April	95	100	90	80	85	90	90
May	90	80	75	60	50	75	95
June	85	90	80	70	75	75	85
July	75	80	75	70	70	70	85
August	70	70	65	60	65	60	65
September ...	65	65	70	50	50	50	70
October	120	110	95	90	80	85	120
November ...	180	175	170	160	165	185	190
December ...	300	320	280	260	255	285	330
Totals	1,590	1,570	1,505	1,360	1,330	1,480	1,665

Solution:

Month.	1945	1946	1947	Sales in £'s.				1951	Seven years' totals.	Arithmetic Mean of seven years.	Seasonal Index.
				1948	1949	1950	1951				
January	200	185	190	175	170	195	215	1,330	190	152	
February	180	170	175	160	150	170	185	1,190	170	136	
March	130	125	140	125	115	140	135	910	130	104	
April	95	100	90	80	85	90	90	630	90	72	
May	90	80	75	60	50	75	95	525	75	60	
June	85	90	80	70	75	75	85	560	80	64	
July	75	80	75	70	70	70	85	525	75	60	
August	70	70	65	60	65	60	65	455	65	52	
September	65	65	70	50	50	50	70	420	60	48	
October	120	110	95	90	80	85	120	700	100	80	
November	180	175	170	160	165	185	190	1,225	175	140	
December	300	320	280	260	255	285	330	2,030	290	232	
Totals	1,590	1,570	1,505	1,360	1,330	1,480	1,665	10,500	1,500	1,200	

Arithmetic Mean:

875 125 100

The procedure to carry out the above working is as follows:

- (1) Arrange the data in columns as in the above solution.
- (2) Total the monthly figures for all the years under review. For example in the above case 1,190 represents the total of all the figures for February for all of the seven years.
- (3) Find the arithmetic average of each of these totals by dividing each total by the number of years. Thus 1,190 divided by 7 gives 170.
- (4) Total these arithmetic averages and divide such total by the number of months (12), to give the arithmetic average per month per year (125).
- (5) The seasonal index is calculated for each month by expressing each monthly average (190, 170, 130, etc.) as a percentage of the average per month per year (125). Thus

$$\frac{190 \times 100}{125} = 152$$

$$\frac{170 \times 100}{125} = 136$$

and so on.

The value of the indexes when so obtained is that they give us a means whereby we can compare the average for a particular month with that of the monthly average per year. January, for instance, with an index of 152 is 52 per cent. higher than the average for the year: May is 40 per cent. below the average for the year and so on.

This monthly average method however is deficient in that it does not take into account the fact that there may be a pronounced trend throughout the data and this will influence the indexes obtained.

The Moving Average Method

The moving average method of calculating a seasonal index tends to remove to some extent the effect of the secular trend and thus is, in that respect, more satisfactory than the monthly average method just described.

The procedure necessary to use the moving average method is extremely tedious and involves the following steps:

- (1) Ascertain the average of the original data for successive twelve month periods on the moving average method. These are set out in column (4) of the table A below.

For instance average of sales for period from January, 1947, to December, 1947 is 100; the average for period from February, 1947, to January, 1948, is 99·08; and so on.

- (2) Pair these averages successively and total them. For instance $100 + 99\cdot08 = 199\cdot08$; $99\cdot08 + 98\cdot33 = 197\cdot41$; $98\cdot33 + 99\cdot08 = 197\cdot41$; and so on. These are inserted in column (5).

- (3) Average each of the totals in column (5). Thus $\frac{199.08}{2} = 99.54$, $\frac{197.41}{2} = 98.71$; etc. These are inserted in column (6).
- (4) Next, express each of the items of original data as a percentage of the respective averages as obtained in step (3) above. Thus, for July, 1947, $\frac{120}{99.54} \times 100 = 120.55$, and for August, 1947, $\frac{110}{98.71} \times 100 = 111.43$, and so on. These are inserted in column (7).
- (5) The percentages in column (7) are then transferred to another table, viz., table B.
- (6) Column (a) in table B carries the totals of the percentages for each month.
- (7) The arithmetic average of each of these totals is next obtained by dividing such totals by the number of years represented in the totals (in this example, 3). These are inserted in column (b).
- (8) The arithmetic average of the total of column (b) is then obtained.
- (9) Each of the averages in column (b) is expressed as a percentage of the average ascertained in step (8). These are the seasonal indexes and are inserted in column (c).

Table A

Year	Month	Sales	12 Months' Average	Total of Paired Averages	Average of Paired Averages	% of Sales each month to average of Paired Averages
1947	January	80				
	February	90				
	March	100				
	April	130				
	May	150				
	June	170				
	July	120	100·00	199·08	99·54	120·55
	August	110	99·08	197·41	98·71	111·43
	September	90	98·33	197·41	98·71	91·17
	October	60	99·08	197·41	98·71	60·78
	November	50	98·33	197·41	98·71	50·65
	December	50	99·08	197·91	98·96	50·52
1948	January	70	100·83	200·83	100·42	69·70
	February	80	100·00	198·33	99·17	80·66
	March	110	98·33	194·99	97·50	112·82
	April	120	96·66	193·32	96·66	124·14
	May	160	96·66	192·49	96·25	166·23
	June	190	95·83	189·99	95·00	200·00
	July	110	94·16	189·16	94·58	116·30
	August	90	95·00	191·66	95·83	93·91
	September	70	96·66	191·66	95·83	73·04
	October	60	95·00	189·16	94·58	63·43
	November	40	94·16	187·49	93·75	42·66
	December	30	93·33	185·83	92·92	32·28
1949	January	80	92·50	184·16	92·08	86·88
	February	100	91·66	182·49	91·25	109·58
	March	90	90·83	180·83	90·42	99·53
	April	110	90·00	180·00	90·00	122·22
	May	150	90·00	180·83	90·42	165·89
	June	180	90·83	181·66	90·83	198·17
	July	100	90·83	180·83	90·42	110·59
	August	80	90·00	179·16	89·58	90·42
	September	60	89·16	177·49	88·75	67·72
	October	60	88·33	177·49	88·75	67·72
	November	50	89·16	179·16	89·58	55·81
	December	30	90·00	180·83	90·42	33·17
1950	January	70	90·83	182·49	91·50	76·50
	February	90	91·66	184·16	92·08	97·74
	March	80	92·50	185·83	92·92	86·09
	April	120	93·33	186·66	93·33	128·57
	May	160	93·33	185·83	92·92	172·19
	June	190	92·50	186·66	93·33	203·57
	July	110	94·16			
	August	90				
	September	70				
	October	60				
	November	40				
	December	50				

Table B

Month.	1947	1948	1949	1950	(a) Total.	(b) Arithmetic Average.	(c) Seasonal Index.
January		69.70	86.88	76.50	233.08	77.69	79.17
February		80.66	109.58	97.74	287.98	95.99	97.82
March		112.82	99.53	86.09	298.44	99.48	101.38
April		124.14	122.22	128.57	374.93	124.94	127.33
May		166.23	165.89	172.19	404.31	134.77	137.35
June		200.00	198.17	203.57	601.74	200.58	204.42
July	120.55	116.30	110.59		347.44	115.81	118.02
August	111.43	93.91	90.42		295.76	98.58	100.46
September	91.17	73.04	67.72		231.93	77.31	78.79
October	60.78	63.43	67.72		191.93	63.97	65.19
November	50.65	42.66	55.81		149.12	49.71	50.66
December	50.52	32.28	33.17		115.97	38.65	39.38
						1177.48	

$$\text{Arithmetic Average} = \frac{1177.48}{12} = 98.12$$

Removal of Seasonal Variations

The following example illustrates how the seasonal index may be applied to the original data to remove season variations. In addition, the secular trend of the corrected data is calculated, so that the percentage deviation from the trend may be computed.

Illustrative Example 6

Sales of a certain commodity during the month of August in each of the nine years were as under:

<i>Year</i>	<i>Sales in £s</i>
1941	1700
1942	1717
1943	1751
1944	1768
1945	1751
1946	1785
1947	1802
1948	1785
1949	1819

Calculation of a seasonal index of sales has shown that the index for August is 85, on the basis of 100 as the arithmetic mean of the monthly averages.

You are required to remove the seasonal variation from the above table of August sales, and then calculate the percentage deviation from the long-term trend of the corrected sales in August 1948.

Solution:

The corrected August sales for each year are found by dividing the actual sales by the seasonal index and multiplying by 100, in each case. These corrected sales figures represent the actual sales after removal of the seasonal variation, and are shown in column (c) of the following table which is also designed to carry the necessary columns for computation of the long-term trend by the use of the formula, $y = mx + b$.

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
Year.	Sales in £'s.	Corrected Sales (y)	x	xy	x ²	Trend Ordinates.	Percentage of items in col. (c) to those in col. (g)	Percentage Deviations from Trend.
1941	1,700	2,000	-4	- 8,000	16	2,012		
1942	1,717	2,020	-3	- 6,060	9	2,028		
1943	1,751	2,060	-2	+ 4,120	4	2,044		
1944	1,768	2,080	-1	- 2,080	1	2,060		
1945	1,751	2,060	0	0	0	2,076		
1946	1,785	2,100	1	2,100	1	2,092		
1947	1,802	2,120	2	4,240	4	2,108		
1948	1,785	2,100	3	6,300	9	2,124	98.87	-1.13
1949	1,819	2,140	4	8,560	16	2,140		
		$\Sigma(y) = 18,680$	—	$\Sigma(xy) = 940$	$\Sigma(x^2) = 60$			

$$b = \frac{\Sigma(y)}{n} = \frac{18,680}{9} = 2,076$$

$$m = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{940}{60} = 16.$$

Substituting these values of m and b in the formula $y = mx + b$, for the year 1941, we have:

$$\begin{aligned} y &= (16 \times - 4) + 2076 \\ &= 2012 \end{aligned}$$

and so on for the remaining years.

These trend values are then inserted in column (g).

The next step is to calculate the percentage of each item in column (c) to its respective trend value in column (g). Thus, in respect of the year 1948 the percentage would be:

$$\frac{2100}{2124} \times 100 = 98.87$$

This could be done for each year if desired. Then in column (i) are shown the percentage deviations from the trend by subtracting 100 from each item in column (h).

Columns (h) and (i) could be used as a possible measure of "irregular fluctuations".

CHAPTER 6

GRAPHIC PRESENTATION OF DATA

INTRODUCTION	99
RULES FOR GRAPHIC PRESENTATION	100
TYPES OF GRAPHIC PRESENTATION	105
1. Line Graphs.								
2. Area Diagrams.								
3. Bar Charts.								
4. Solid Diagrams.								
5. Cartograms.								
6. Pictograms.								

INTRODUCTION

A graph is a symbolic diagram representing or expressing a set or sets of statistical data. It is a visual form of presentation.

The terms, *graph*, *diagram* and *chart* are more or less synonymous and are words of a general nature indicating that data have been depicted in some representative form calculated to make them readily assimilable to the mind. These terms, therefore, are used more or less interchangeably throughout this chapter, except that occasionally certain types of graphs bear particular names, e.g. in the case of a Pie Chart, the term Chart is used and not diagram or graph. However, an attempt at classification into graphs, diagrams and charts would prove fruitless.

As indicated, the great advantage of a *graphic presentation* as compared with a *tabular presentation* is that the former depicts the data in a way which facilitates absorption by the mind. This is an invaluable attribute from the point of view of two classes of persons, viz.:

- (a) the person who is unskilled in interpreting tabulated data, and,
- (b) the person who requires a speedy summing up of a set of statistical facts.

As regards the former case, increasing use is being made of graphic presentations of data by company directors in their reports to members. Illustrations of these are given throughout this chapter and there is little doubt that they have a great value in making the average member better informed of the activities of his company.

The validity of the statement made in (b) above, is almost beyond dispute. Not only does a graph provide a mental picture far easier to comprehend than a set of tabulated figures but also it is quite often the only form in which the data may be interpreted, e.g. the chart of a plotted course used by the navigator.

Graphic presentation of data has its uses in most forms of intellectual activity. Instances of these are, the temperature chart used by the doctor, the seismogram by the geologist, the calibration curve by the engineer, the concentration curve by the chemist, the amplitude curve by the physicist and so on. As far as accounting data are concerned it will be found that they may be represented in a great variety of ways.

RULES FOR GRAPHIC PRESENTATION*

It is essential that the preparation of graphs should be carried out in accordance with rules and principles which are more or less uniformly recognized. Set out below are the main rules which should be observed in preparing graphs.

1. (a) A zero line for the vertical scale, should, if possible, appear on the diagram and the scale selected should provide for this. (See Figure 1 (a).)
- (b) If, having regard to the size of the numerical quantities or values to be placed on the vertical scale, the showing of a zero line would not normally be convenient, then the vertical scale should be broken if possible to permit of the zero line being shown. (See Figure 1 (b).)
- (c) The ruling of the zero line should be heavier than that of the other co-ordinate lines. (See Figure 1 (c).)
- (d) Particularly is this so where the zero line appears in the middle of the diagram. (See Figure 1 (d).)
2. (a) In the case of a graph representing percentages, it is desirable that the 100 per cent. line be heavier than the other co-ordinate lines. (See Figure 2 (a).)
- (b) Particularly is this so where the 100 per cent. line appears in the middle of the diagram. (See Figure 2 (b).)
3. (a) If more co-ordinate lines than necessary are used, the diagram may tend to become confusing. (See Figure 3 (a).)
- (b) Therefore, it is preferable to use only such number of co-ordinate lines as will enable the diagram to be easily read. (See Figure 3 (b).)

* These rules are based on those produced by the Joint Committee on Standards for Graphic Presentation, America.

No comparable rules have been produced in this country although it is believed that the British Institute of Management is considering the matter.

4. Greater emphasis should be given to the line representing the graph of the data than to the lines forming the ruling of the diagram. (See Figure 4.)
5. Where possible, it is desirable to indicate on the diagram the individual points representing the items of the data. (See Figure 5.)
6. It is often desirable to accompany the diagram by a tabular statement showing the data of which the graph is a representation. (See Figure 6.)
7. The title of the diagram should be as brief, yet clear and complete as possible. (See Figure 6.)

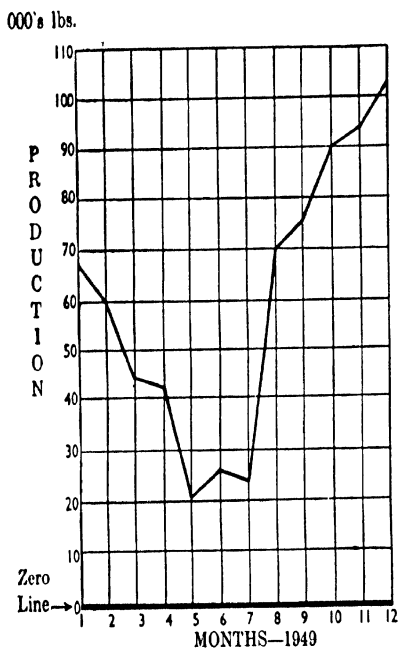


Fig. 1 (a).

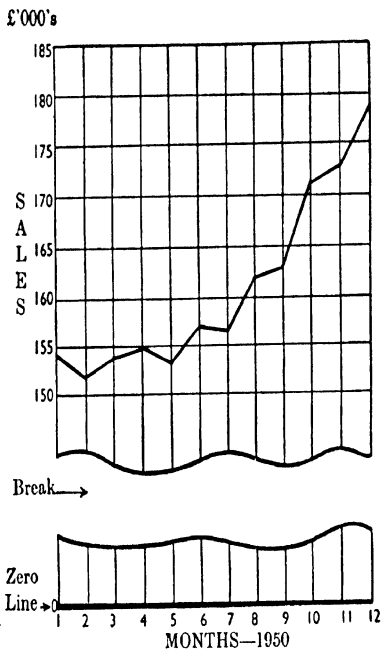


Fig. 1 (b).

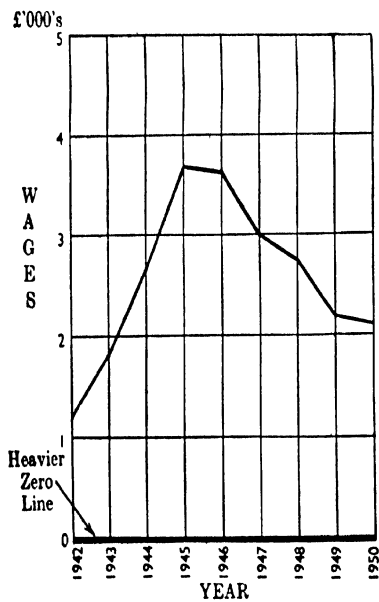


Fig. 1 (c).

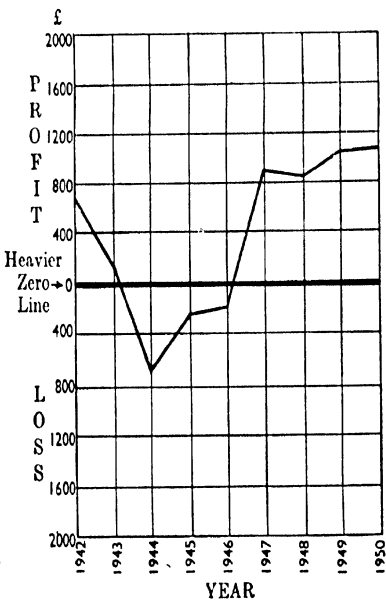
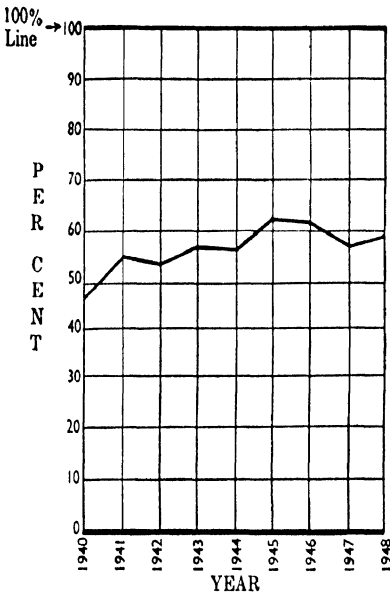
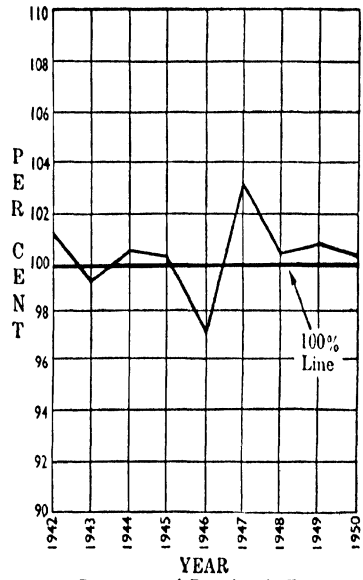


Fig. 1 (d).



Percentage of Gross Profit to Sales.

Fig. 2 (a).



Percentage of Proprietor's Funds to Fixed Assets.

Fig. 2 (b).

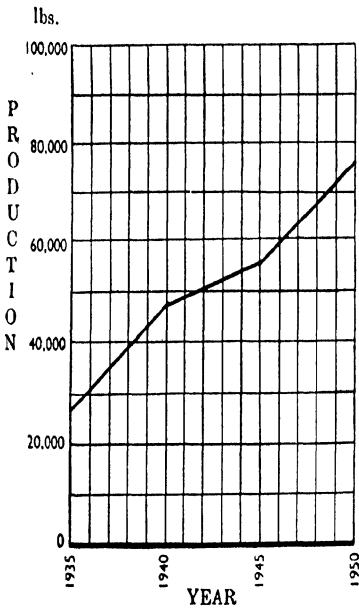


Fig. 3 (a).

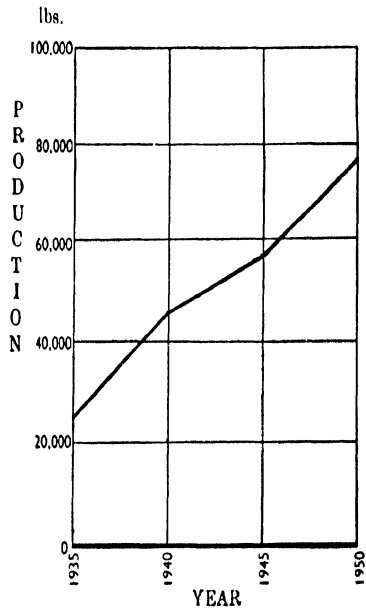


Fig. 3 (b).

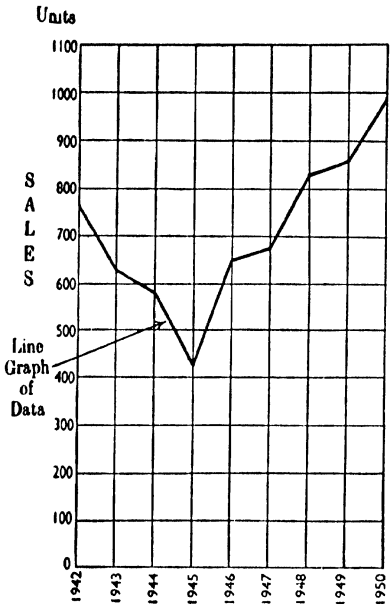


Fig. 4.

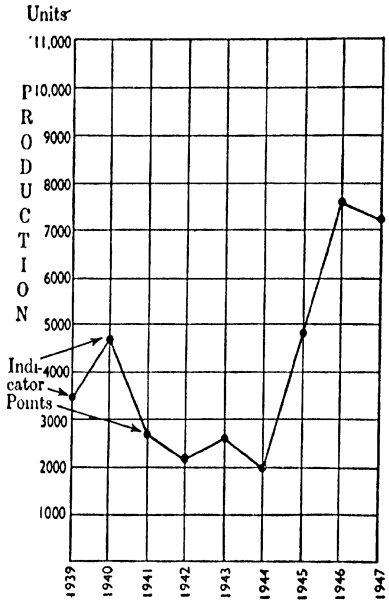
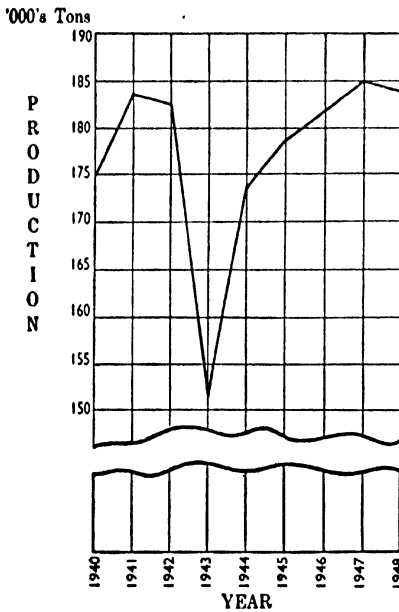


Fig. 5.

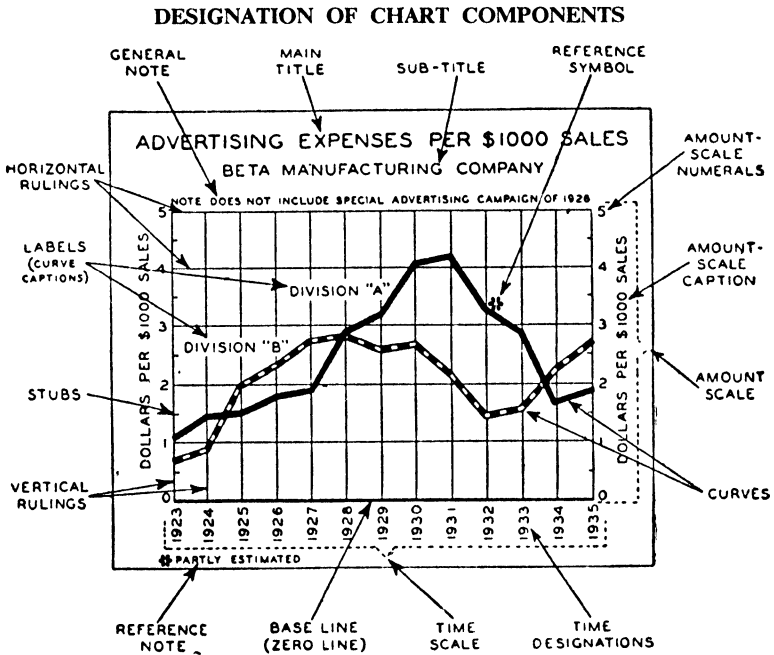


Year	Production Tons
1940	175,000
1941	184,000
1942	183,000
1943	151,000
1944	174,000
1945	179,000
1946	182,000
1947	185,000
1948	184,000

Factory Production to nearest thousand tons.
Partial loss of plant by fire in 1943.

Fig. 6.

Set out below is a chart showing the designation of the various components of a graph, which has been reproduced from the Standard Rules on Graphic Presentation.



TYPES OF GRAPHIC PRESENTATION

The various types of graphs may be most conveniently classified as follows:

1. Line graphs.
2. Area diagrams.
3. Bar Charts.
4. Solid diagrams.
5. Cartograms.
6. Pictograms.

Each of these types will now be briefly considered.

1. Line Graph

The purpose of a Line Graph is to indicate variations in the data either by means of a straight or a curved line. The rules for preparing this type of graph have been dealt with above and it now remains to consider the different types of line graphs that may be used.

The most satisfactory classification of line graphs is one based upon the type of scale ruling used. These are—

- (a) arithmetical rulings,
- (b) logarithmic and semi-logarithmic rulings,
- (c) miscellaneous rulings.

(a) Arithmetic Rulings

An arithmetic scale is one marked out by divisions, between each two of which there is the same distance.

Two points should be noted in connection with this type of ruling, viz.:

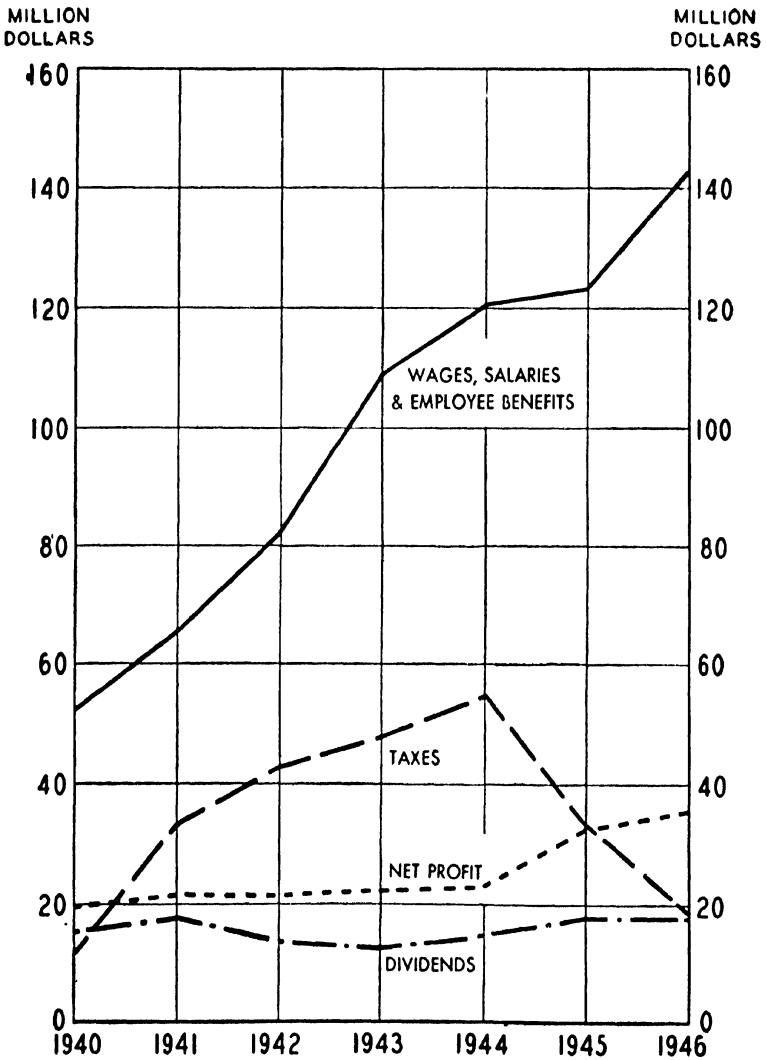
- (i) if an arithmetic progression* is plotted it will result in a straight line;
- (ii) as equal distances on the scale represent equal amounts, the graph will reveal the actual extent of *absolute* increases or decreases.

This type of graph wherein an arithmetic or natural scale is employed is by far the most commonly used. On pages 107 and 108 are illustrated two line graphs employing arithmetic scales, taken from the Annual Report of the American Eastman Kodak Company

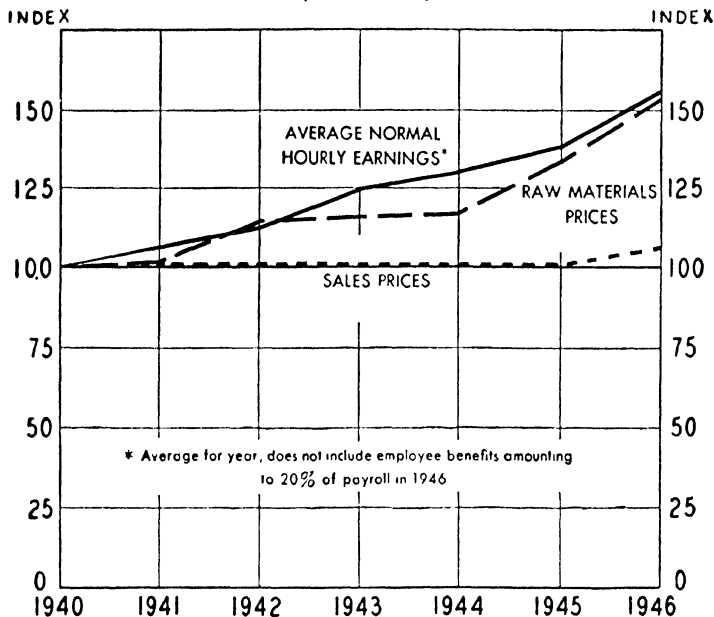
* An arithmetic progression is one wherein each term differs from the one preceding it by an equal amount, e.g. 2, 9, 16, 23, 30, 37 . . . is an arithmetic progression because the difference between each pair of terms is 7, or put another way, there is a constant difference of 7 between the successive values of the series.

GRAPHIC HIGHLIGHTS OF SIGNIFICANT YEARS

WAGES, SALARIES & EMPLOYEE BENEFITS, TAXES, NET PROFIT, AND DIVIDENDS



SALES PRICES AND MAJOR COSTS (PHOTOGRAPHIC PRODUCTS) (1940=100)



(b) Logarithmic and Semi-Logarithmic Rulings

If a logarithmic ruling is used on both the x and the y axis, the resultant graph is known as a *Logarithmic Chart*. If logarithmic ruling is used on one axis only, it is known as a *Semi-Logarithmic Chart*. These forms of ruling are used where it is desired to compare *percentage* rather than *absolute* changes. The reason for this may be illustrated as follows:

The log. of 4 is 0.60206
and ,, ,, 12,, 1.07918

and the difference 0.47712

Also log. of 7 is 0.84510
and ,, ,, 21,, 1.32222

and the difference 0.47712

It will be noticed that between the numbers 4 and 12 there is the same rate of change or the same percentage change as there is between 7 and 21, and the difference between their logarithms is the same. Therefore, if in a series, the logarithms of the values are plotted instead of the original values themselves, any constant differences, i.e. rises and falls, will equal constant percentage changes.

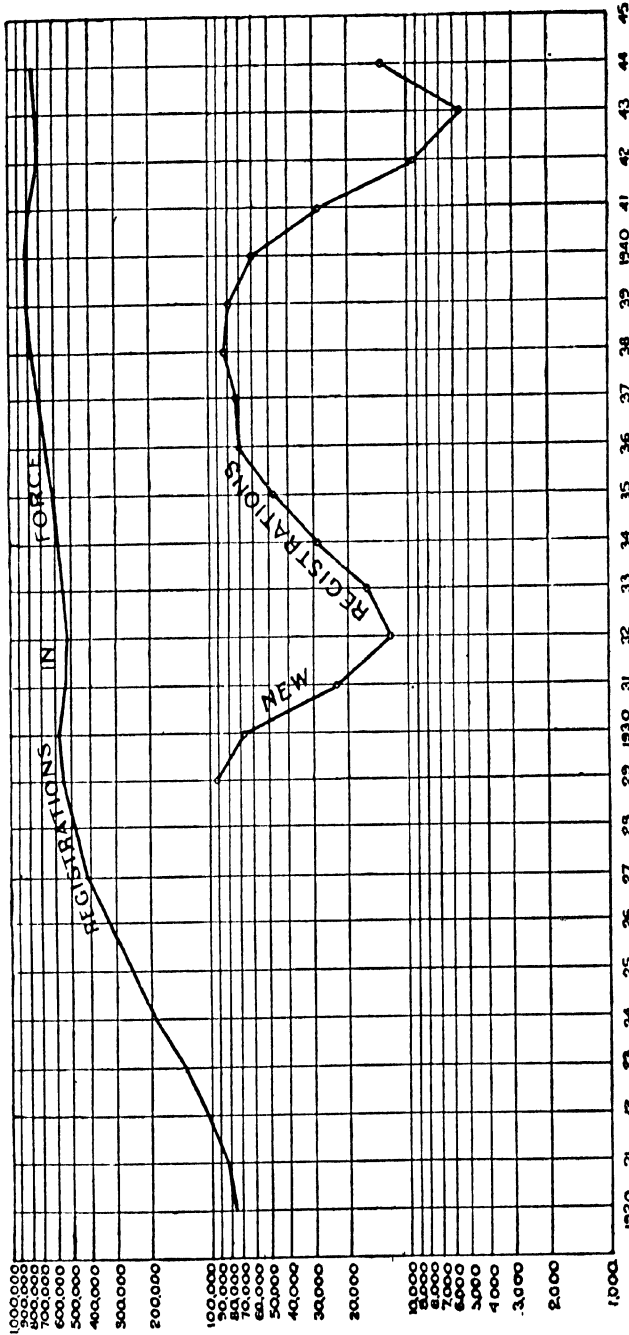
There are two ways in which this type of graph (sometimes called Ratio Graph) may be constructed. The first is where the values of the items are converted to logarithms and these logarithms plotted on an ordinary or arithmetical scale. Thus, if it is desired to plot the number 21 according to a logarithmic scale, its logarithm, viz. 1.3222, will be ascertained from the tables and 1.3222 plotted on an arithmetical scale. This, as can be imagined, may be very tedious where many numbers are involved. The second and by far the most common way is to use specially ruled paper (generally called ratio-ruled paper).

Here the distances between the horizontal lines are according to a logarithmic scale and the great advantage of this method is that the *actual* numbers or values may be plotted on the Vertical Scale.

On page 110 is an illustration of a ratio graph extracted from the Official Year Book of the Commonwealth of Australia (No. 36).

RATIO GRAPH

MOTOR VEHICLE REGISTRATION—AUSTRALIA, 1920 TO 1944.



EXPLANATION.—This is a ratio graph, the vertical scale being logarithmic, and the curves rise and fall according to the rate of increase or decrease. Actual numbers are indicated by the scale at the side of the graph. The graphs show for all motor vehicles other than motor cycles the registrations in force at 30th June each year and the number of new registrations during the year.

Features of Ratio Graphs

The following significant features associated with ratio graphs require to be carefully noted:

- (i) If the vertical distance between a pair of horizontal lines is the same as that between another pair of horizontal lines, then the proportional change between each pair is the same. For instance, in the graph on page 110 the vertical distance between 2,000 and 4,000 is the same as that between 10,000 and 20,000 and there is the same proportional change between 2,000 and 4,000 (viz. 2) as there is between 10,000 and 20,000. Thus if a geometric progression* were plotted on a logarithmic chart it would form a straight line. This is so because the logarithms of the terms of a geometric progression constitute an arithmetic progression.
- (ii) Two or more different sets of data may be plotted on the same graph (e.g. costs and profits). They can thus be compared by means of the slopes they exhibit. If the slopes are more or less parallel, for instance, then as between the two sets of data there is approximately the same proportionate rate of change. Further, if one part of a curve is steeper than another it means that the former part is changing at a rate more rapid than the latter.
- (iii) If it is desired to compare on the one graph two series which are widely separated as to values (e.g. profits may be in the £10,000 to £20,000 group but costs say in the £100,000 to £150,000 group) then plotting the series on a ratio chart will have the effect of bringing them closer together and thus facilitate their comparison.
- (iv) It has been stated elsewhere that *Time* is generally placed on the horizontal or x axis, and therefore in a Time series an arithmetic scale is generally used on this axis and a logarithmic ruling used on the vertical or y axis. The resulting graph will then be a semi-logarithmic chart.
- (v) There is no base line and zero cannot be shown on a ratio graph.

Set out hereunder is an extract from the Standard Rules for Graphic Presentation dealing with Ratio Graphs.

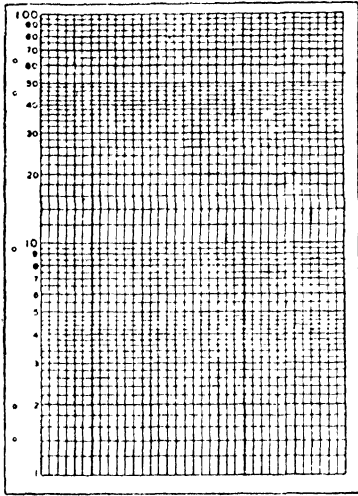
RATE-OF-CHANGE CHARTS

A. Definition

The rate-of-change chart ("ratio" or "semi-logarithmic" chart) is a type used for picturing the percentage or relative change

* A geometric progression is a series wherein there is the same proportionate change between each pair of adjacent items in the series. Thus, 3, 9, 27, 81 . . . is a geometric progression since $\frac{9}{3} = 3$, $\frac{27}{9} = 3$, $\frac{81}{27} = 3$. . .

in values of a series over a period of time rather than the change in absolute amounts as shown by the arithmetic chart.



Two-cycle semi-log paper scale
from 1 to 100

1. The picture of rate of change is achieved through the use of logarithms. Rate-of-change curves can be constructed either by plotting the logarithms of the values on an arithmetic scale or by plotting the actual values on a logarithmic scale. The latter is the more usual procedure.

2. The effective use of rate-of-change charts requires an appreciation of their limitations as well as their possibilities.

B. When to use Rate-of-change Charts.

1. When the interest is in relative movement of a time series and not in the differences between amounts.

2. When it is desired to compare the relative movements of several time series.
3. When the readers are likely to be familiar with this form of chart.
4. When the usual arithmetic chart would present a misleading picture of movement.
5. For occasions when there are no minus figures included in the time series.

Note: If it is desired to present a complete picture of both rate of change and amount of change the data can be presented on companion charts, one with a logarithmic amount scale and the other with the usual arithmetic scale.

CONSTRUCTION

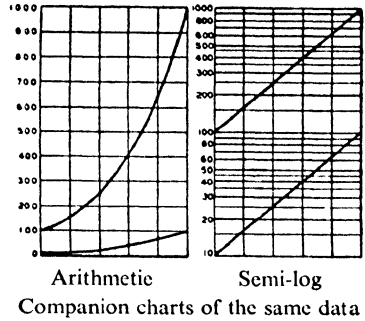
1. Layout and Design

In general, the principles and procedures required in the preparation of rate-of-change charts are the same as for arithmetic charts.

2. Grids

Rate-of-change charts are normally prepared on commercially printed papers. On these forms the horizontal rulings are spaced

logarithmically and the vertical rulings are spaced arithmetically. The scale range on "single cycle" paper includes only a single power of 10 (1-10, 10-100, etc.) but paper is available which provides for several cycles (1-100, 1-1,000, etc.). These grids are generally closely ruled horizontally to assist in plotting.



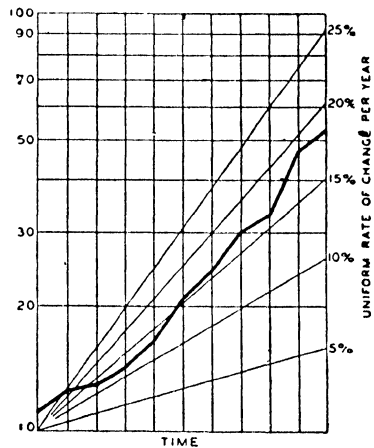
Companion charts of the same data

Where it is desired to simplify rate-of-change charts and commercial grids do not prove entirely satisfactory, special grids, having the desired characteristics, can readily be laid out by the "diagonal scale method" using a commercial paper as a guide.

3. Scale Selection

Logarithmic amount scales should be so selected that the curves will be well placed on the grid. As there is no zero line to serve as a base for comparing trends, considerations of the zero line are not applicable to rate-of-change charts.

In rate-of-change charts, it is often helpful to provide a secondary scale indicating uniform rates of change. Such scales are constructed by means of straight diagonal lines radiating from some point of origin (as shown in the illustration).



Secondary scale of uniform rates of change

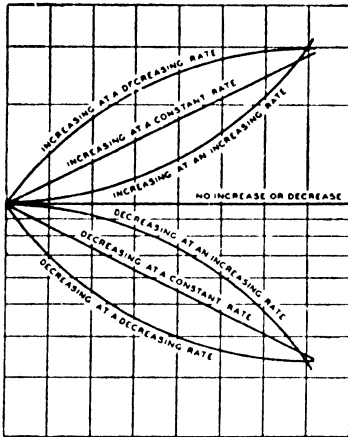
Multiple amount scales are more appropriate for rate-of-change charts than for arithmetic charts because in the former the movement of the curves is compared and not their position relative to a base. The procedures relative to the zero line as presented on pages 102 to 104 do not apply, but otherwise the procedures are similar for both arithmetic and rate-of-change charts.

Note: In order to take full advantage of the scale range, the vertical scale numerals usually printed on a chart sheet may be multiplied by any constant factor but integers should be chosen so that the scale subdivisions will not indicate inconvenient fractions.

4. Scale Designations (See 7 below).

5. Curves

The plotting on rate-of-change charts requires considerable care because of the peculiar character of the logarithmic spacing. Where special grids are prepared without intermediate rulings, it is desirable to use a logarithmic plotting scale which may easily be made from printed commercial paper of the proper dimensions.



Meaning of curve shapes on
Rate-of-change charts

In general, rate-of-change charts call for simple lines connecting the points of value. Columns or surfaces, of course, should not be used to indicate values on a rate-of-change chart. Columns and surfaces may be used on an arithmetic chart to indicate changes in ratios, however.

6. Curve Designations (See 7 below.)

7. Chart Designations

In general, the principles and procedures in regard to scale, curve and chart designations are the same as for arithmetic charts.

(c) Miscellaneous Rulings of Line Graphs

There is a number of miscellaneous forms in which line graphs may be presented. The most common are—

- (i) *Silhouette Charts*. Here the positive or negative deviations from a base line are plotted, the plotted points joined to give the line graph and then the area between the line graph and the base line is filled in.

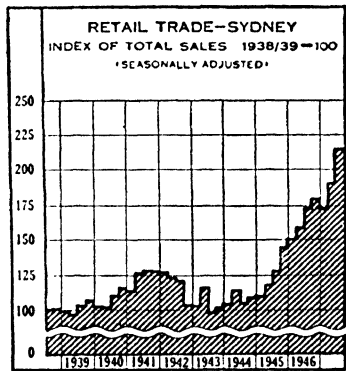
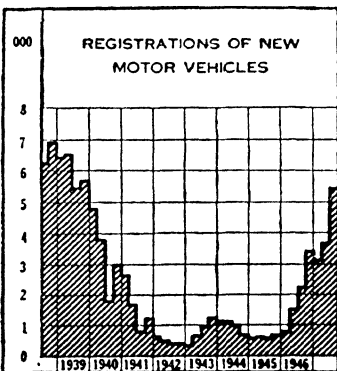
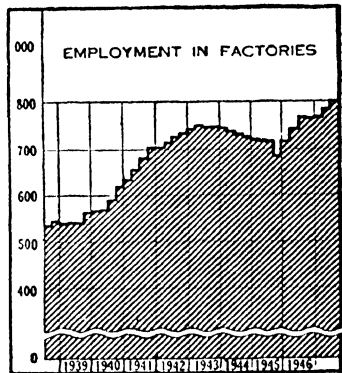
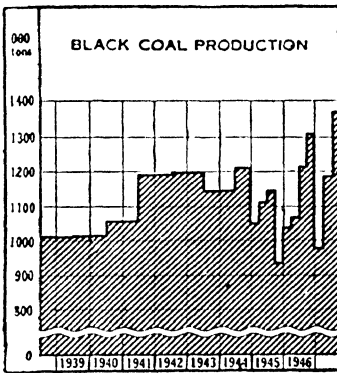
An example of a Silhouette Chart is given on page 76 in Chapter 5, whilst another example is given below, taken from the Recovery Guide of the Economic Co-operation Administration.

- (ii) *Histograms and Historigrams*. In the case of a histogram rectangles are formed out of the data so that the frequency of each series is represented by the height of each rectangle and the class interval is represented by the width of the rectangle. When so drawn the area enclosed by the sum of all the rectangles may be filled in as is the case with the illustrations below taken from the Statistical Bulletin of the Commonwealth Bank of Australia.

An histogram is the term used for this type of diagram drawn to fit a time series. Thus, the only difference between an histogram and an histogram is that the former refers to a *frequency* series and the latter to a *time* series.

The histogram is sometimes called a rectangular frequency polygon.

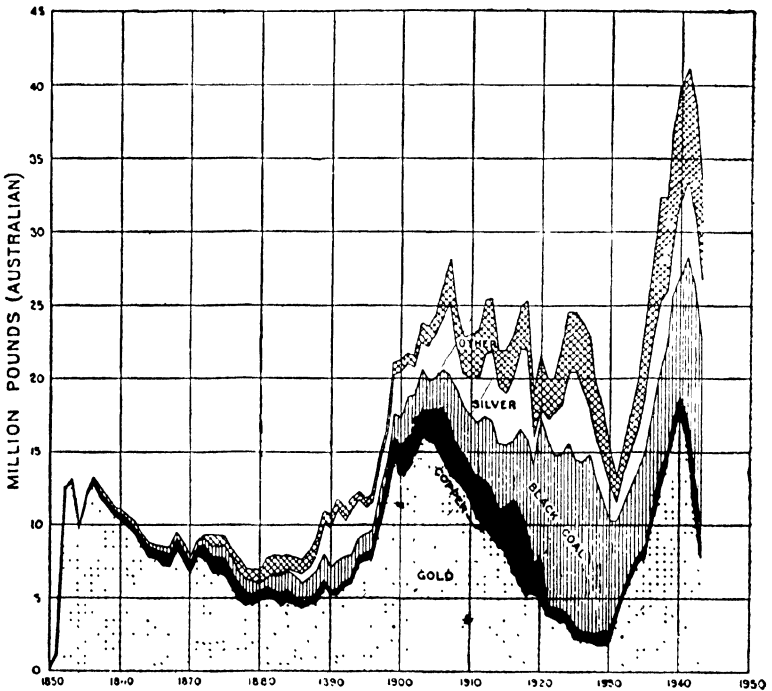
Historigrams



- (iii) *Band Chart*. This is a particularly valuable form of line graph since it shows not only the graph of the total of the various component parts of the whole but also the individual graphs of the parts themselves. (See the explanation at the foot of the Band Chart illustrated below.) This has been taken from the Official Year Book of the Commonwealth of Australia.

BAND CHART

VALUE OF PRINCIPAL MINERALS PRODUCED—AUSTRALIA, 1850 TO 1943



EXPLANATION.—The upper curve represents the total value of mineral production while the vertical distances between the curves represent the value of production of each mineral.

2. Area Diagrams

Area diagrams may take different forms but they all involve the same idea, viz. to contrast different quantities by relating such quantities to areas. For this purpose squares, rectangles, circles and other geometrical figures are used, but perhaps the most common type of area diagram is the Pie Chart. By permission of H.M. Stationery Office a series of Pie Charts are given on page 145. These are reproduced from *Recovery Record, The Story of Marshall Aid in Britain*.

The method of construction of a Pie Chart to illustrate the distribution of a company's costs would be as follows:

- (i) Tabulate the data so as to show the respective percentage of each item to the total.

Thus:

	1943
(a) Cost of Goods	61 %
(b) Employees Received	16½ %
(c) Rent and Light, Advertising and Maintenance, etc.	6½ %
(d) Taxation	13 %
(e) Shareholders Received	2 %
(f) Reserves	1 %
	100 %

- (ii) Assume the full circle to represent 100 per cent.
- (iii) Since at the centre of the circle there are four right angles or 360° , then each per cent. represents $\frac{360^\circ}{100}$, i.e. 3.6° .
- (iv) Draw in arcs of the circle such that each respective arc bears an angle at the centre corresponding to the number of degrees arrived at by multiplying the per cent. of the item by 3.6° . Thus item (e) above represents an angle at the centre of 2×3.6 degrees, i.e. 7.2° .
- (v) The sectors should run clockwise with the largest first so that a start will be made by drawing a vertical line from the centre of the circle to the position of twelve o'clock on the circumference of the circle.

The chart illustrated on page 119 is based on one issued by the Central Office of Information which gave in pictorial form the details of the 1951/2 Budget. The original was in colour, revenue in dark green and expenditure in red. The use of bold colourings in such pictograms can be *extremely effective*. Alongside the legend were small pictures to bring alive the figures. For example, the word "defence" was illustrated by a soldier, a warship, and an aircraft, and "Alcohol" by a foaming tankard of ale.

3. Bar Charts

Bar Charts may be classified as—

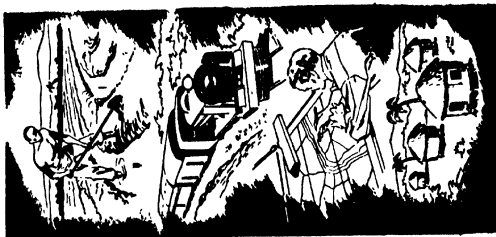
- (a) *Vertical Bar Charts*, where the bars are vertical; or
 (b) *Horizontal Bar Charts*, where the bars run horizontally across the graph.

DEVELOPMENT OF OVERSEAS TERRITORIES

ECA OPERATIONS**③ TECHNICAL ASSISTANCE**

Only through a substantial increase in technical and administrative skills will these underdeveloped areas succeed in raising productivity and improving living standards. For this reason the metropolises, with ECA aid, are carrying out technical assistance programs to go hand in hand with the material aid provided.

By June 30, 1951, ECA had obligated \$914 thousand for technical assistance projects in the overseas territories of Marshall Plan countries the bulk of the money for assistance in transportation, agriculture and land development, and an aerial survey.

**AGRICULTURE AND LAND DEVELOPMENT—\$272,000**

This aid will enable the territories to cope with several major obstacles to increased agricultural productivity, e.g., inadequate soil analysis and insufficient knowledge of proper soil management; water control through irrigation, flood control, and tapping of new sources of water; and, plant diseases, pests (especially locusts), and animal diseases. A number of other ECA technical assistance projects relate to improving methods of cultivation, cattle and sheep raising, and raw foodstuffs processing.

TRANSPORTATION—\$100,000

ECA is providing technicians to the Belgian Congo to advise in the establishment of maintenance shops for highway construction and maintenance equipment, and for engineering surveys to establish railway routes.

AERIAL SURVEY—\$500,000

The largest obligation for a technical assistance project in the overseas territories is \$500,000 for an aerial survey of certain portions of the Portuguese territories of Angola and Mozambique. This survey is part of a larger geographic survey for the discovery of new sources of strategic materials. Ground surveys will follow the aerial mapping project.

OTHER FIELDS OF TECHNICAL ASSISTANCE—\$42,000

In addition to the types of assistance described above, ECA has approved technical assistance projects in the overseas territories relating to health improvement, housing, industry, and public administration.

THE NATIONAL BUDGET

1951-52

REVENUE

This is how the Government collects each £ of Revenue

DIRECT TAXES

11/2

INCOME TAX and SURTAX
(£1,753 million)

8/3

DEATH DUTIES PROFITS TAX, etc.
(£493 million)

2/4

EVERYTHING ELSE
e.g. stamp duties, motor taxes, etc.
(£125 million)

7d.

INDIRECT TAXES

7/10

PURCHASE TAX
(£310 million)

1/6

ALCOHOL
(£368 million)

1/9

TOBACCO
(£400 million)

2/10

ENTERTAINMENTS, Betting, etc.
(£74 million)

4d.

EVERYTHING ELSE

e.g. petrol tax and other Customs & Excise duties
(£292 million)

1/5

NON-TAX REVENUE

e.g. trading services, surplus war stores, etc.
(£221 million)

1/0

EXPENDITURE

This is how the Government spends each £ that it collects

DEFENCE AND NATIONAL DEBT

9/6

DEFENCE
(£1,488 million)

7/0

NATIONAL DEBT
e.g. interest on War Loans, Savings Certificates, etc.
(£535 million)

2/6

SOCIAL SERVICES AND FOOD SUBSIDIES

7/11

EDUCATION, HOUSING and other local services
(£455 million)

2/2

HEALTH
(£398 million)

1/11

PENSIONS, ALLOWANCES, etc.
(£396 million)

1/11

FOOD SUBSIDIES
(£410 million)

1/11

EVERYTHING ELSE

Works, Buildings 2d. Agriculture & Fisheries 3d. Colonial & Foreign 5d. Ministry of Supply (Civil) 2d. Employment Services 3d. Broadcasting 1d. Research and Development 1d. Civil Aviation 1d. Rest 11d. Surplus 2d.
(£554 million)

2/7

TOTAL £4,236 million

£1.0.0.

TOTAL £4,236 million

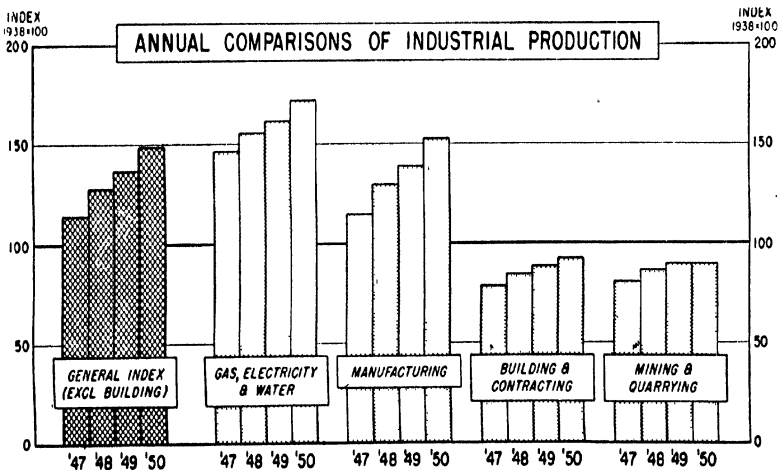
Those diagrams below and on pages 122 and 123 are vertical bar charts while that on page 121 is a horizontal bar chart. The horizontal bar chart has been taken from the pamphlet on Energy by Dr. Bronowski published by the Bureau of Current Affairs.

Either vertical bar charts or horizontal bar charts may be—

- (i) *Simple*, i.e. consisting of single bars representing one unit only.
- (ii) *Compound or multiple*, i.e., a grouping of bars representing different units so as to throw emphasis upon comparisons. The illustration below has been taken from the Report of the Economic Co-operation Administration.
- (iii) *Component*, i.e. the whole bar is divided into its component parts and thus shows the total as well as the individual variations. The illustration on page 122 has been reproduced from the pamphlet on Social Insurance published by the Bureau of Current Affairs.

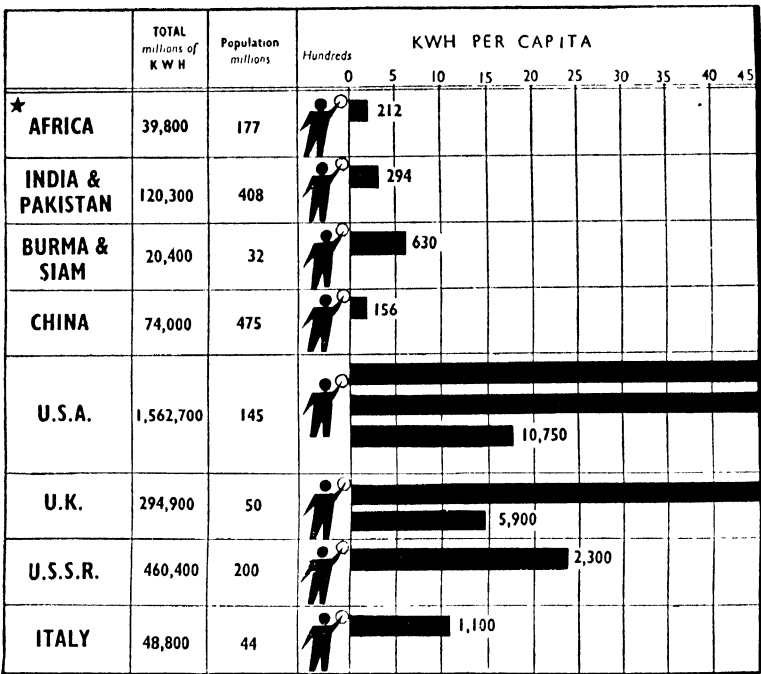
Bar Charts may be representations of absolute values, i.e. the actual items themselves, or percentages. An illustration of a percentage bar chart is given on page 123, reprinted here from the Monthly Summary of Business Conditions in Southern California, produced by the Research Department of the Security-First National Bank of Los Angeles.

COMPOUND VERTICAL BAR CHART
UNITED KINGDOM



CONSUMPTION OF ENERGY—1948

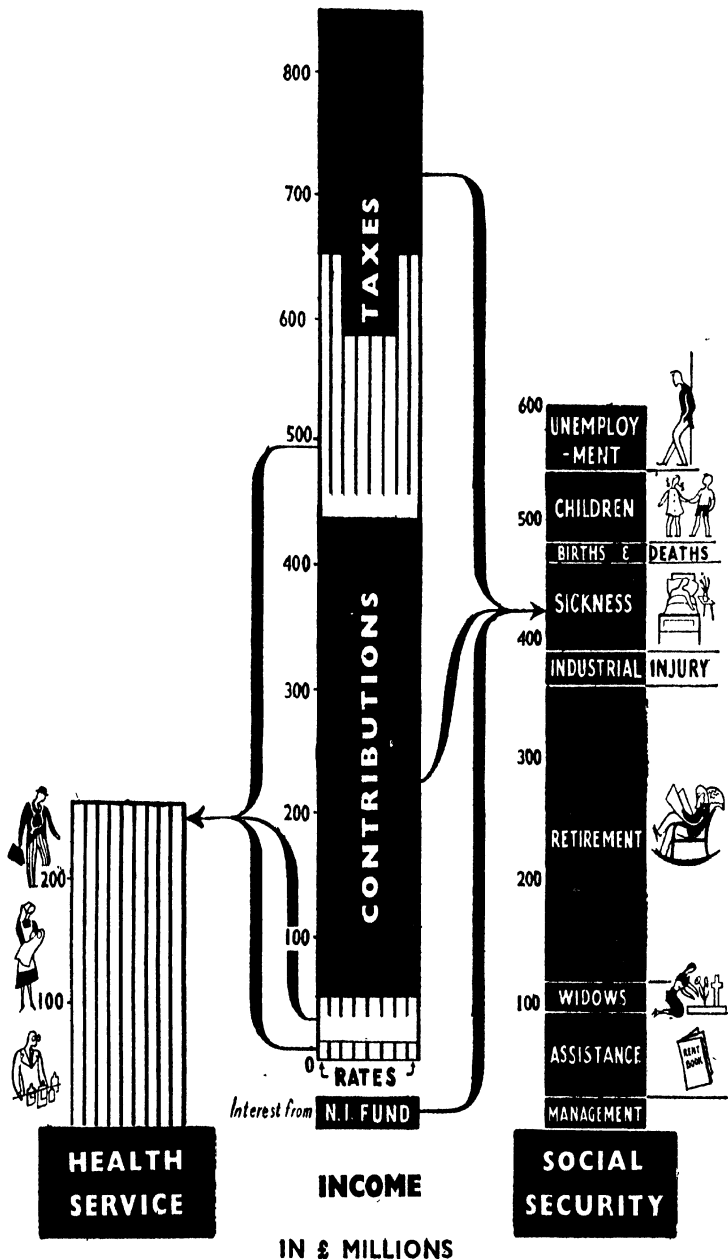
The estimated total consumption of energy for production purposes expressed in Kilowatt-hours



* Excludes South Africa

From the Guyol Report, 1949

HOW THE MONEY COMES & GOES



A further variant of the bar chart is reproduced below by permission of the Glacier Metal Co. Ltd.

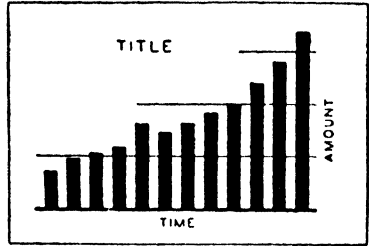
THE GLACIER METAL CO. LTD.
FINANCIAL YEAR 1947/8

SURPLUS REVENUE		HOW USED	
NET TRADING PROFIT	£288,498	EXTRA STOCKS AND WORK IN PROGRESS	£137,056
RESERVES RELEASED (Depreciation £40,000) (Contingencies 10,000)	£50,000	EXTRA DEBTORS (owing to us by customers)	£114,572
INCREASE IN CREDITORS (Inland Revenue £48,185) (Sundry Creditors 39,832)	£88,017	EXTRA PLANT	£44,663
BALANCE OF CASH USED UP DURING YEAR	£30,617	TAXATION	£147,481
	<u>£457,132</u>	PREFERENCE DIVIDEND	£5,544
		ORDINARY DIVIDEND	£7,816
			<u>£457,132</u>

Set out hereunder is an extract from the Standard Rules for Graphic Presentation dealing with Bar Charts. These charts are referred to in the rules as Column Charts.

TIME-SERIES COLUMN CHARTS

A. *Definition.*—Column charts are graphic presentations wherein numerical values are represented by the length of vertical bars or columns.



Simple column chart

B. *The Column Chart is particularly effective:*

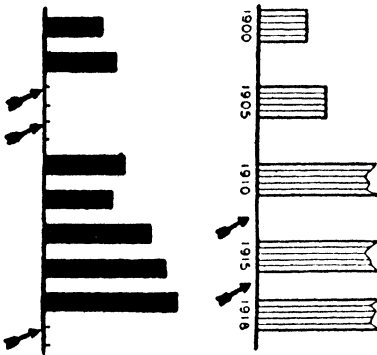
1. To emphasize comparisons of amounts in a single time series.
2. For popular presentation.
3. To show components for a relatively few totals.
4. To picture "period" data as against "point" data.
5. For showing a range of values or deviations from a normal or bogey.

C. *The Column Chart is not the best form:*

1. For comparing several times series.
2. For time series over an extended period with many plottings.

Note: The column chart, while simple to read, is relatively difficult to design effectively. The proper layout of columns requires special care and good judgment.

CONSTRUCTION



Indicating irregularities in time scale

1. *Layout and Design.*—A chart consisting of a few columns should generally be higher than wide; for more than a few columns a wider-than-high chart is preferable.

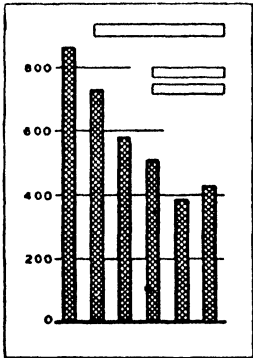
2. *Grids.*—The field or grid used for column charts may be a completely ruled coordinate surface. Usually, however, it is not necessary to indicate all the rulings which would normally be shown on a line chart. A

complete grid outline is usually not required. The columns themselves generally make vertical rulings unnecessary. Moreover, fewer horizontal rulings may be needed since column charts are

more generally used for popular presentation than are line charts. Often horizontal rulings may be incomplete, being extended through only that portion of the field occupied by the columns.

3. *Scale Selection*.—In column charts the interest is generally in a comparison between amounts as of different dates. These amounts are proportionate to the height of the columns. This means that the zero line, when it is the principal line of reference, should always be included in a column chart. It follows, too, that the amount scales should not be broken, but made continuous from the reference line. While normally the full length of the column should be shown, when it represents an abnormally large value the column may be broken at the top and the amount indicated.

Columns should be spaced according to their proper position on the time scale. When time intervals between values are not equal, columns should be spaced accordingly.



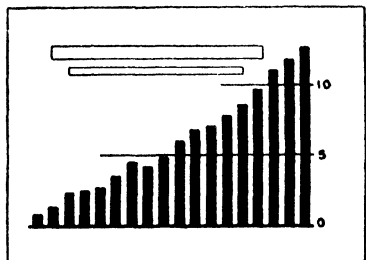
Few columns—higher than wide

4. *Scale Designations*.—Placing of scale numerals and captions on column charts is less conventionalized than on line charts. As the grid rulings are often incomplete, the vertical scale values generally are placed on the side where the rulings are complete. (For example, if the tallest columns are at the right, the scale designations may be shown on the right-hand side only.)

Time Scale Designations are normally centred under the columns, reading horizontally; in column charts for popular presentation either or both amount and time

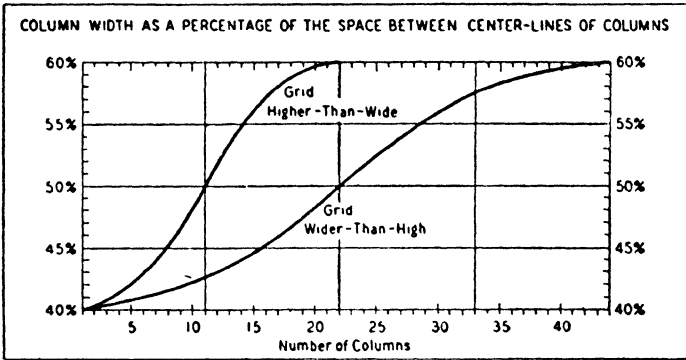
designations may be placed above the columns (see illustration above).

5. *Columns*.—The effective appearance of a column chart requires special care in the design of the columns. When there are only a few columns they should be narrower than the white space between; when there are many columns the reverse should be true.

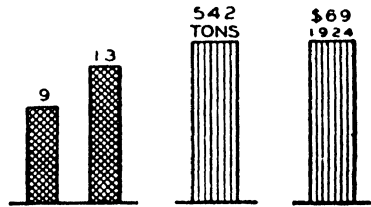


Many columns—wider than high

Note: An empirical relationship between column and space is presented in the chart below, based on an actual test of charts of root-two proportions and various numbers of columns, one set wider-than-high and another higher-than-wide.



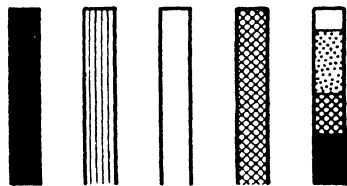
To space columns equally along the time scale, divide the available horizontal space into twice as many spaces as there are to be columns. Then centre the columns on every other division mark beginning with the first from either end.



Variations of location of scale designations

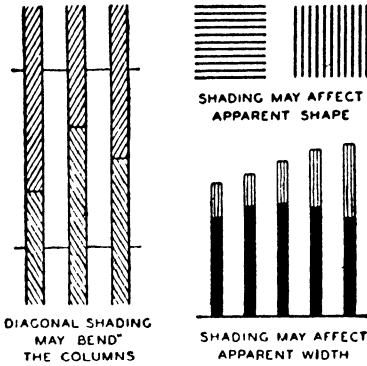
Lightly shaded columns should be slightly wider than solid black columns to appear the same width.

As the number of columns increases the chart loses its character as a column chart and becomes, in effect, a shaded surface chart. For a chart of root-two proportions, wider-than-high, the range of columns that can be effectively shown is from about 8 to 40 columns; for a chart higher-than-wide, from 1 to about 16 columns.



Examples of shading for columns

The weight and spacing of the lines and dots of the shading should be determined from a consideration of the size of the areas to be shaded and the relation between the sizes of the original and finished chart.



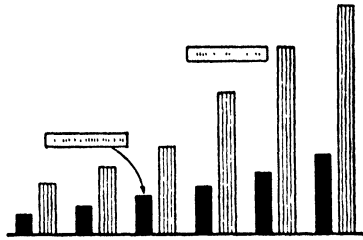
Effects of improper use of shading

Specific column designs or shadings are recommended as follows:

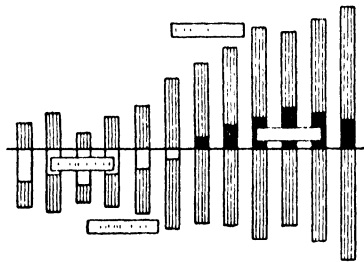
- (a) Black (solid) for general use for narrow columns. However, a series of long narrow columns filled in solid may cause an unpleasant optical effect. In segmented column charts, black is good for the bottom segments if they are not too large.
- (b) Vertical Line Shading is recommended for general use as pleasing in appearance and easy to construct.
- (c) Diagonal Line Shading is useful only in small segments as optical illusion results if any appreciable length of column is shaded with this design, as illustrated above.
- (d) Horizontal Line Shading has limited usefulness and is not generally recommended.
- (e) Cross-hatch Shading (diagonal) is recommended in place of black for wide columns. Cross-hatch shading made by crossing vertical and horizontal lines is not recommended.
- (f) Dotted Shading (pebbled or stippled) is sometimes effective for columns of medium width and particularly for small segments for charts in which a third or fourth distinguishing shading is needed.
- (g) Hollow columns, if distinctly wider or narrower than the space between and outlined with a heavy line.

Columns may present undesirable optical illusions unless slight correctives are applied. A white or lightly shaded segment on top of a column may appear to spread unless the column outline is tapered about the width of a line; a black segment may appear more narrow than the rest of the column unless it is widened about the width of a line; a tall column may appear to be thinner in the middle unless the lines are bowed out slightly.

6. *Column Designations.*—It is generally more difficult to label segmented or grouped columns than curves because the columns themselves take up so much more of the space. Segment labels should be placed across several columns if practicable. However the space about labels should be reduced as much as possible and too much contrast with the tone of the column avoided so as not to distort the impression of the relative lengths of the columns and segments. Where labels cannot be placed on the columns, arrows may be used. A key or legend should be used only when impracticable to label directly.



Designations for grouped columns



Designations for "deviation" columns

7. *Column Chart Designations.*—Column chart titles can often be placed most effectively according to the distribution of the columns rather than in a fixed position at the top (the usual case with line charts).

4. Solid Diagrams

These consist of pictorial representations of geometric forms such as spheres, cubes, etc. The purpose is to throw into relief comparisons of different units by contrasting the volumes of the solids representing the units. This means that a person making an inspection of solid diagrams must generally be expert at judging their volumes. For this reason this form of graphic presentation is generally little used.

5. Cartograms

The cartogram is simply a map representing a spatial distribution. For instance, a map of Britain may be prepared in such a way as to indicate the different areas wherein sales of a company's product are made. Indications may be by dots or squares or shading

or cross-hatching, all varied in proportion so as to show differences in numbers or densities. This form of graphic presentation is particularly important to the Sales executive.

6. Pictograms.

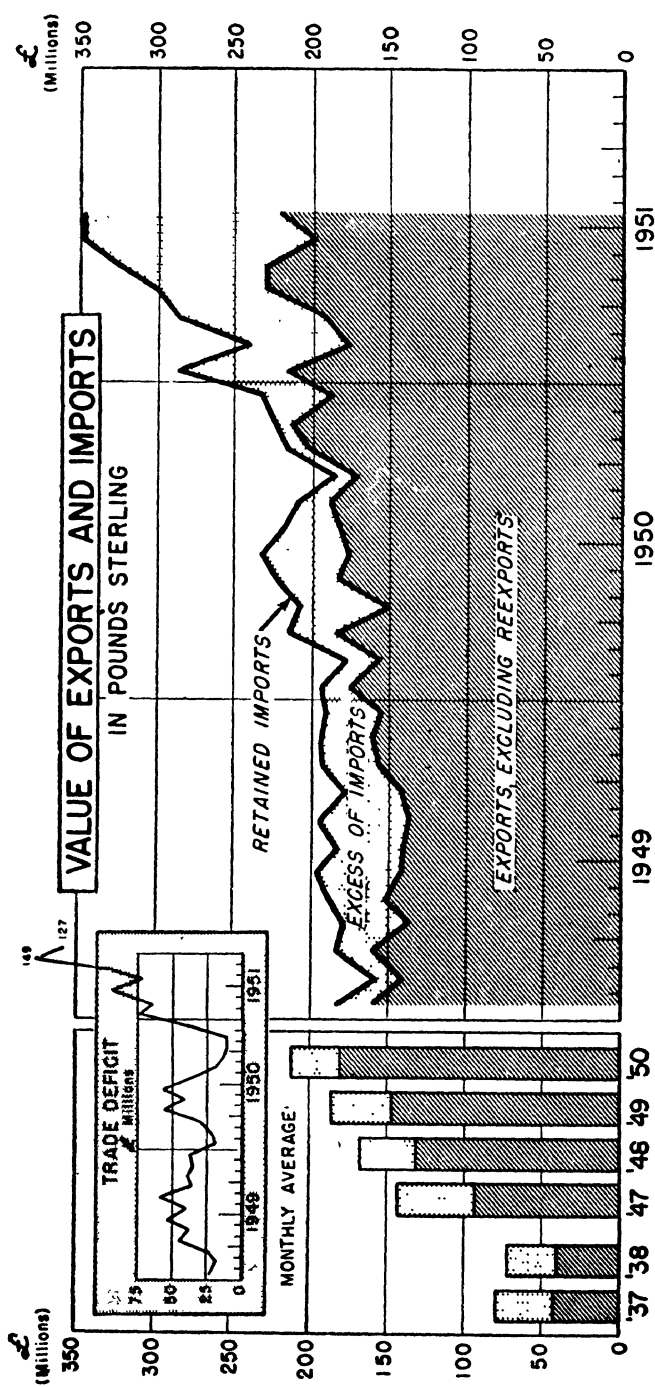
These are diagrams of infinite variety, consisting of drawings of articles designed to emphasize pictorially differences or variations in the units compared. Increasing use is being made of the pictogram by directors in their annual reports to the members of the Company. It is a form of graphic presentation extremely easy to comprehend and further, the impression it leaves on the mind usually lingers for some time after it is withdrawn from sight.

Illustrations of the pictogram are given in the following pages.

The pictogram on page 133 was taken from the report of the Economic Co-operation Administration, whilst that on page 132 is reproduced from "Marshall Aid in Britain" by permission of H.M. Stationery Office.

UNITED KINGDOM

FOREIGN TRADE



BRITAIN'S WAR LOSSES



£1,500 m.

Damage to property
(e.g. 1 house in 3 damaged)



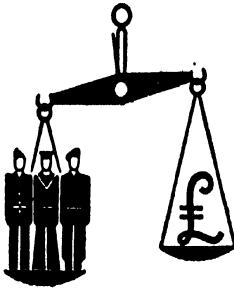
£700 m.

Loss of shipping and cargoes
(half of the pre-war British
Merchant Fleet was sunk)



£900 m.

Wastage and deterioration of
equipment (five years gap in
replacement and re-equipment)



£2,900 m.

War-time debts to overseas
countries



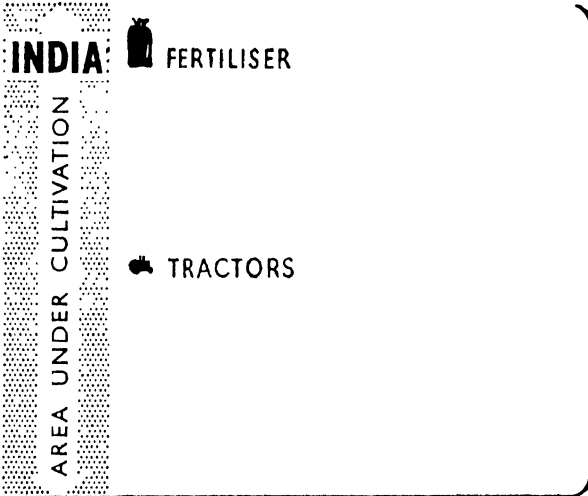

£1,100 m.

Sale of overseas assets


£200 m.

Fall of gold and dollar reserves

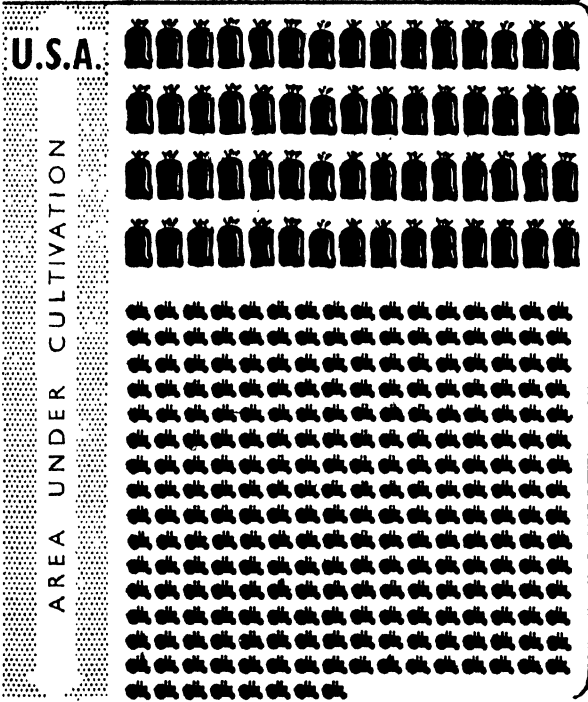
TOTAL LOSS $\left(\frac{1}{4}\right)$ **OF BRITAIN'S NATIONAL WEALTH**
(or £7,300 million)

Wheat yield per acre one half of that in U.S.A.



Cotton yield per acre one fifth of that in U.S.A.





WHAT MACHINES CAN DO

The area under cultivation in the United States is only about one-sixth larger than in India, yet the United States uses over sixty times as much fertiliser — more than 13,000,000 tons a year, compared with only 200,000 in India. The

CHAPTER 7

SAMPLING: ACCURACY AND APPROXIMATION: INDEX NUMBERS

SAMPLING	134
Principles of Selection.									
Methods of Selecting the Sample.									
ACCURACY AND APPROXIMATION	136
Methods of Approximation.									
Error.									
Biased Error.									
Unbiased Error.									
Illustrative Example 1.									
INDEX NUMBERS	141
The Construction of Index Numbers.									
Illustrative Example 2.									
Illustrative Example 3.									
Illustrative Example 4.									
Illustrative Example 5.									

SAMPLING

In practice, it is not always possible or desirable to tabulate, classify and analyse the whole of the data relating to a particular problem. This may be so because the volume of data is so great that the time and expense involved in its complete analysis would not be justifiable or profitable; or again, it may not be possible to collect the whole of the data.

In cases such as this, the fact that the data is not handled in its entirety does not preclude its subjection to statistical analysis. By the process of *sampling*, representative portions of the entire mass of data are selected and examined. The representative portions are known as *Samples* and the entire mass of data from which the samples are drawn is known as the *Universe*, or *Population*.

The theory relating to sampling is based on several laws or principles, the most important of which are:

- (a) the Law of Statistical Regularity, and
- (b) the Principle of Large Numbers.

The Law of Statistical Regularity is to the effect that if a reasonably large sample is selected from a population consisting of a large number of items, then such sample will, in general, possess characteristics little different from those of the population from which

it was drawn. The Principle of Large Numbers states that, although abnormalities occur or exist in a large population, the probability is that the abnormalities in one direction will be offset by the abnormalities in the opposite direction so that the average of the population will remain unaffected.

Principles of Selection

Certain conditions should be observed to ensure that the sample is as representative of the population as possible. These are:

- (a) The data should, if possible, be grouped and the samples selected from each of the groups. For instance, if it is desired to select from among 3,000 employees, samples of times taken for the carrying out of a particular task, the samples would not be representative if they were all related to those carried out by young people. The employees, in such a case, should first of all be grouped into convenient age classes and samples drawn from each of them.
- (b) Bias or prejudice should be completely avoided in selecting the samples.
- (c) Every item of the population should have an equal chance of being selected as a sample.
- (d) Every item of the population should be subject to the same conditions. For instance, following on the example in (a) above, the samples would not be representative of the population if 1,000 of the employees carried out the operation using old machinery and the other 2,000 carried out the operation using modern and improved types of machines.
- (e) The greater the number of samples selected the greater will be the accuracy achieved in the result.

Methods of Selecting the Sample

Random selection of samples may be effected in many ways but the two most common are (a) the Lottery Method, and, (b) the Array Method.

By the Lottery Method a draw is made as is done in the case of a lottery in order to choose the winner. The great advantage of this method is that if all the items comprising the population are well mixed, each item has the chance of being selected in the draw. An illustration of how this method is used is as follows. Assume that a large selling organization desires to test or sample the reaction of the public to one of its products and for that purpose decides to send representatives to question residents of 6 out of the 30 suburbs of the city throughout which the product is sold. The choosing of the 6 suburbs may be done by the lottery method,

i.e., all of the names of the 30 suburbs are shuffled together and 6 drawn at random.

In the case of the Array Method, the data is arrayed in ascending or descending order and every n th item selected, n being small enough to give a sufficient number of items in relation to the total to form an adequate sample. Actually it is not always necessary for data to be arrayed to apply this method. For example, a production manager may test a mass-produced product by examining every, say, seventh article produced.

ACCURACY AND APPROXIMATION

Although accounting data must be recorded accurately in the books of account and records, the accountant often deals with figures for statistical purposes in cases wherein absolute accuracy is not necessary. In that event approximations are made so as to facilitate calculation. Sometimes, too, approximations are made in financial statements themselves, e.g. by the elimination of shillings and pence from the figures stated in published revenue statements and balance sheets.

Generally speaking, approximations should only be made where they will assist in computation and yield a result which is substantially accurate, i.e. so close to the true result that the difference may be treated as negligible. It is important also, to take into consideration the nature of the item, e.g. whereas the output of a coal mine may be taken to the nearest ton, the output of a gold mine would be taken to the nearest ounce.

Methods of Approximation

The following are the methods sometimes used to approximate numbers to those capable of easy manipulation:

- (1) *Approximation to the nearest whole number below.*

Here the procedure is to discard the last digits and substitute for each a zero.

In the following case, "the nearest whole number" taken was the "nearest hundred". It could have been taken to the nearest thousand or to the nearest ten.

Thus:

Actual Numbers	Approximations
1792	1700
641	600
1809	1800
1728	1700
1847	1800
429	400
98	—
<hr style="width: 50%; margin: 0 auto;"/> 8344	<hr style="width: 50%; margin: 0 auto;"/> 8000

Difference = 344

This method has the disadvantage of producing what is called a Biased Error, for an explanation of which see *post*.

Percentages under this method, would be approximated as follows—

76.289% would be stated as 76%
 32.769% „ „ „ „ 32%

It is usual, however, to approximate percentages correct to two places of decimals, so that 76.289% would become 76.29% and 32.769% would become 32.77%.

- (2) *Approximation to the nearest whole number above.*

In this case the numbers in (1) above would be approximated as follows:

Actual Numbers	Approximations
1792	1800
641	700
1809	1900
1728	1800
1847	1900
429	500
98	100
8344	8700

Difference = 356.

Like the previous method this one also produces Biased Error.

- (3) *Approximation to the nearest whole number whether above or below.*

Under this method, the actual numbers appearing in the previous methods would be approximated as follows:

Actual Numbers	Approximations
1792	1800
641	600
1809	1800
1728	1700
1847	1800
429	400
98	100
8344	8200

Difference = 144.

Here, the difference is much less than is the case with the other methods. This method results in what is known as an Unbiased Error.

Error

The term *Error* as used in statistics refers to the difference between a computed (estimated) value and the true value. Error does not mean “mistake” as used in ordinary parlance.

There are two ways of expressing the degree of error, viz., (a) as an absolute measure, (b) as a relative measure.

Absolute Error is the difference between the approximated figure and the actual figure, e.g. if 9425 had been approximated to 9500, the absolute error is 75.

Relative Error is the ratio of the absolute error to the approximation, so that in the above case the relative error is:

$$\frac{75}{9500} = 0.0079$$

This relative error may be expressed as a percentage error by multiplying by 100. So:

$$\begin{aligned} \text{Percentage Error} &= 0.0079 \times 100 \\ &= 0.79\%. \end{aligned}$$

Classes of Error

The two classes into which errors may be divided are:

- (a) Biased Errors.
- (b) Unbiased Errors.

Biased Errors are those which result when the approximations have all been made to the nearest whole number below, or have all been made to the nearest whole number above. That is, the approximations have all been made on the "same side" of the actual numbers.

Taking the figures used in methods (1) and (2) on pages 136 and 137, the comparison would be as follows:

Actual Numbers	Approximations to nearest whole number below	Approximations to nearest whole number above
1792	1700	1800
641	600	700
1809	1800	1900
1728	1700	1800
1847	1800	1900
429	400	500
98	—	100
<hr style="width: 50%; margin: 0 auto;"/> 8344	<hr style="width: 50%; margin: 0 auto;"/> 8000	<hr style="width: 50%; margin: 0 auto;"/> 8700
Absolute Biased Error	344	356

It will be seen therefore that the biased error is a cumulative error and the greater the data the greater will be the error.

Unbiased Error results where the numbers have been approximated to the nearest whole number whether above or below. In a lengthy list of numbers, the effect of this will be for the differences to compensate each other and thus reduce the absolute error.

For example:

Actual Numbers	Approximations
72	70
98	100
41	40
67	70
29	30
33	30
21	20
17	20
88	90
81	80
547	550

Absolute Unbiased Error = 3.

Estimation of Errors in Totals

Where a number of figures has been approximated and added it is possible to estimate the absolute or relative error whether biased or unbiased.

Biased Errors

In the case of approximations made, for example, to the nearest ten below, the greatest possible error in any one item is 9, but on the average it would be $\frac{9}{2}$ i.e., 4.5. This is called the average error. To estimate the absolute error in the sum of a number of items which has been so approximated, this average error is multiplied by the number of items. In Illustrative Example 1 the estimated absolute biased error is therefore:

$$4.5 \times 11 = 49.5$$

To estimate the relative biased error, divide the above result by the estimated total. In this case the estimated relative biased error would be:

$$\frac{49.5}{7130} = 0.0069$$

which expressed as a percentage error would be:

$$0.0069 \times 100 = 0.69\%.$$

Unbiased Errors

In the case of approximations made, for example, to the nearest ten (whether above or below) the greatest possible error in any one item is 5, but on the average would be 2.5.

To estimate the absolute unbiased error in the sum of a number of items which has been so approximated, multiply the average error by the square root of the number of items. Thus in Illustrative Example 1 the estimated absolute unbiased error would be:

$$\begin{aligned} 2.5 \times \sqrt{11} &= 2.5 \times (\pm 3.31) \\ &= \pm 8.275. \end{aligned}$$

To express this as a relative error, divide by the estimated total. The estimated relative unbiased error in this case is therefore:

$$\frac{\pm 8.275}{7200} = \pm 0.00115$$

which expressed as a percentage error would be:

$$\pm 0.00115 \times 100 = \pm 0.115\%$$

Illustrative Example 1

In respect of the following set of numbers you are required to add them by approximating them to the:

- (a) nearest 10 below
 (b) nearest 10 above or below

and estimate in each case the biased and unbiased absolute error and the biased and unbiased relative error.

946
 762
 375
 210
 496
 782
 817
 375
 937
 824
 657

Solution:

Actual Numbers	Approximations to nearest 10 below	Approximations to nearest 10
946	940	950
762	760	760
375	370	380
210	210	210
496	490	500
782	780	780
817	810	820
375	370	380
937	930	940
824	820	820
657	650	660
	<hr/> 7130 <hr/>	<hr/> 7200 <hr/>

Estimated Biased Error

$$\text{Absolute} = 4.5 \times 11 = 49.5$$

$$\text{Relative} = \frac{49.5}{7130} = 0.0069$$

Estimated Unbiased Error

$$\text{Absolute} = 2.5 \times \sqrt{11} = \pm 8.275$$

$$\text{Relative} = \frac{\pm 8.275}{7200} = \pm 0.00115$$

It is emphasized that in the above illustrations only a small number of items was taken to illustrate the principles involved. It is not intended that the addition of such a small number of items be carried out by the methods of approximation as it would

be quicker in such case to add the actual items. The principles outlined however may be used with a large mass of numbers to facilitate addition where 100 per cent. accuracy is not required.

Similar principles may be used for approximating in connection with the multiplication, division, subtraction and averaging of numbers, but a detailed description of the respective procedures is not within the scope of this text.

INDEX NUMBERS

Index numbers are statistical devices prepared for the purpose of measuring changes in data at different points of time. They are devices which serve as a means of comparative measurement of the data or groups of data which they represent and thus the principle underlying their preparation is that of reducing all of the factors to be compared to a common denominator.

Originally, index numbers were used to study movements in Price Levels and the Purchasing Power of Money. To-day, however, their use has been extended to cover not only the above types of movements but also changes in Profits, Costs, Production, Sales, Interest Rates, etc. They are frequently used by individual enterprises to measure changes in the above types of data applicable to the particular organizations.

By means of index numbers the average price of a number of articles in, say, 1935, may be compared with that for the same number and type of articles in, say, 1950. Generally speaking, it is the purpose of index numbers to supply an indication of the trend of some group of data in this way. Many different series of index numbers are published in the official publications of the Board of Trade, the Ministry of Labour, the Government Monthly Digest of Statistics. These include index numbers for retail prices, wholesale prices, basic materials and foodstuffs, wage rates, employment, production, export prices, etc.

In order to simplify the explanation of how index numbers are constructed it will be assumed that we are concerned with those relating to prices.

The Calculation of Index Numbers.

The study of Index Numbers is such a complex one and is of such considerable extent that a detailed description of even some of the procedures is quite beyond the scope of a text on Advanced Accounting. It is considered, however, that the accountant should be familiar with one or two of the more simple and generally used methods and a brief outline of these is now given. It should be noted that in all of the illustrations taken, the prices of only

small number of commodities are used for the purpose of simplifying the description of the particular method. Normally, a far greater number of products or commodities is used in constructing the indexes. In addition it should be noted that in the illustrations which follow, index numbers have been calculated, using the arithmetic average (for the sake of simplicity). This average is not always satisfactory and the geometric mean is to be generally preferred (see Chapter 2 under Geometric Mean and Price Indexes).

The simplest method of calculating index numbers is that based on the *aggregate of actual prices*. The following example illustrates the procedure.

Illustrative Example 2

Calculate a composite index number for the commodities indicated, for the year 1949, using as the base year the figures relating to 1939.

Commodity	Unit	Prices per Unit	
		1939	1949
Milk	Pint	3d.	5d.
Bread	Pound	6d.	7d.
Butter	Pound	1s. 7d.	2s. 0d.

Solution:

Commodity	Prices per Unit	
	1939	1949
Milk	3d.	5d.
Bread	6d.	7d.
Butter	1s. 7d.	2s. 0d.
	<hr/> 2s. 4d.	<hr/> 3s. 0d.

If 1939 is taken as the base year and represents 100 per cent., then the composite index number for 1949 will be:

$$\frac{3s. 0d. \times 100}{2s. 4d.} = 128.57.$$

That is, the prices in the aggregate in respect of the above commodities, have increased 28.57% as compared with those for 1939. This method suffers from a number of disadvantages, the chief of which is that a commodity with a large price will dominate the index.

In order to overcome this disadvantage, the method based on the *average of relative prices* may be used.

Illustrative Example 3

Using the same data as in Illustrative Example 2, calculate a composite index number based on the simple average of relative prices.

Solution:

Commodity	1939		1949	
	Price	Index	Price	Index
Milk	3d.	100	5d.	166.66
Bread	6d.	100	7d.	116.66
Butter	1s. 7d.	100	2s. 0d.	126.31
		300		409.63
Index Number ==		300	==	409.63
		3		3
	==	100	==	136.54

Under this method, price relatives are obtained for each of the commodities. A price relative is the percentage of the price of a commodity for the period under review (in this case 1949) to that applicable to the base period (in this case 1939). Thus the price relative for milk would be calculated as follows:

$$\frac{5 \times 100}{3} = 166.66.$$

The price relatives when so obtained are then averaged to find the index number.

Both of the methods described in Illustrative Examples 2 and 3 produce what are generally called Simple Index Numbers, i.e. Index Numbers which do not take into consideration the fact that different commodities may vary considerably in importance. In other words the index numbers are unweighted and assume that each of the items is of equal importance.

Where, however, it is necessary to calculate index numbers for commodities which vary in importance then a method should be used which gives due consideration or "weight" to each of the items. It is thus necessary to calculate a *weighted index number*. The weights to be assigned to each item may be the actual quantity say, produced or sold, etc., or may be arbitrary numbers estimated to give to the items the respective degrees of importance which they should assume in the calculations.

Illustrative Example 4

From the following data calculate a composite index number for the undermentioned commodities using a weighted aggregate of actual prices, taking the base year as 1939.

Commodity	Weights	Prices	
		1939	1949
A	2000	£1	£1 10 -
B	500	2	3 - -
C	3000	3	2 - -

Solution:

Commodity	Prices			Weights	Price times	
	1939	1949			1939	1949
A	£1 - -	£1 10 -	-	2000	2000	3000
B	2 - -	3 - -	-	500	1000	1500
C	3 - -	2 - -	-	3000	9000	6000
					<u>12000</u>	<u>10500</u>

$$\begin{aligned} \text{Index Number} &= \frac{10500 \times 100}{12000} \\ &= 87.5. \end{aligned}$$

The following example shows the method of computing a weighted index number basing it on price relatives.

Illustrative Example 5

Assuming the detail in Illustrative Example 4, calculate a composite index number based on the weighted average of relative prices.

Solution:

Commodity	Prices			Weights	Price Relatives	Products
	1939	1949				
A	£1 - -	£1 10 -	-	2000	150	300000
B	2 - -	3 - -	-	500	150	75000
C	3 - -	2 - -	-	3000	66.666	199998
				<u>5500</u>		<u>574998</u>

$$\begin{aligned} \text{Index Number} &= \frac{574998}{5500} \\ &= 104.54 \end{aligned}$$

It will be noticed that although the same data was used both in this and the former example, the index number in Illustrative Example 5 is 104.54 whilst that for Illustrative Example 4 is 87.5. This difference is of course due to the particular way in which the weights were applied to the price relatives in the latter example.

In the former example we weighted the *actual prices* whereas in the latter we weighted the *price relatives*.

APPENDIX C

Periodic Payment to Extinguish £1 and Interest

$$\text{Formula: } S = 1 - \left(1 + \frac{r}{100}\right)^{-n}$$

where S = periodic payment applicable to £1.
 r = rate per centum per annum.
 n = number of years.

Note: The values of S are given by the numbers to six places of decimals in the body of the table.

n	r = 2%	3%	4%	5%	6%	7%
1	1.020000	1.030000	1.040000	1.050000	1.060000	1.070000
2	0.515050	0.522611	0.530196	0.537805	0.545437	0.553092
3	0.346755	0.353530	0.360349	0.367209	0.374110	0.381052
4	0.262624	0.269027	0.275490	0.282012	0.288592	0.295228
5	0.212158	0.218355	0.224627	0.230975	0.237396	0.243891
6	0.178526	0.184598	0.190762	0.197018	0.203363	0.209796
7	0.154512	0.160506	0.166610	0.172820	0.179135	0.185553
8	0.136510	0.142456	0.148528	0.154722	0.161036	0.167468
9	0.122515	0.128434	0.134493	0.140690	0.147022	0.153487
10	0.111327	0.117231	0.123291	0.129505	0.135868	0.142378
11	0.102178	0.108078	0.114149	0.120389	0.126793	0.133357
12	0.094560	0.100462	0.106552	0.112825	0.119277	0.125902
13	0.088118	0.094030	0.100144	0.106456	0.112960	0.119651
14	0.082602	0.088526	0.094669	0.101024	0.107585	0.114345
15	0.077826	0.083767	0.089941	0.096342	0.102963	0.109795
16	0.073650	0.079611	0.085820	0.092270	0.098952	0.105858
17	0.069970	0.075953	0.082199	0.088700	0.095445	0.102425
18	0.066702	0.072709	0.078993	0.085546	0.092357	0.099413
19	0.063782	0.069814	0.076139	0.082745	0.089621	0.096753
20	0.061157	0.067216	0.073582	0.080243	0.087185	0.094393
21	0.058785	0.064872	0.071280	0.077996	0.085005	0.092289
22	0.056631	0.062747	0.069199	0.075971	0.083046	0.090406
23	0.054668	0.060814	0.067309	0.074137	0.081279	0.088714
24	0.052871	0.059047	0.065587	0.072471	0.079679	0.087189
25	0.051220	0.057428	0.064012	0.070953	0.078227	0.085811

INDEX

A

- Accuracy and Approximation**, 136 *et seq.*
- Arbitrary Origin**, 27
- Area Diagrams**, 116
- Arithmetic Average**, 25 *et seq.*
- Arithmetic Progression**, 106
- Arrangement of the Data**, 16
- Assumed Mean**, 27
- Average**, 24
 - arithmetic, 25 *et seq.*
 - advantages of, 30
 - and trend, 29
 - arbitrary origin, 27
 - assumed mean, 27
 - class intervals, use of, 29
 - disadvantages of, 30
 - guessed mean, 27
 - short method of calculating, 27
 - simple arithmetic, 25
 - weighted, 28
- deviation**, *see* **Dispersion**
- features of, 25
- geometric mean**, 38
 - advantages of, 39
 - and price indexes, 38
 - disadvantages of, 39
 - how calculated, 38
- harmonic mean**, 40
- median**, 31 *et seq.*
 - advantages of, 34
 - calculation of, 32
 - disadvantages of, 35
 - interpolation, where necessary, 33
- mode**, 35 *et seq.*
 - advantages of, 37
 - calculation of, 35
 - disadvantages of, 37
 - types of, 25

B

- Band Charts**, 116
- Bar Charts**, 117 *et seq.*
- Biased Error**, 138

C

- Cartograms**, 129
- Category Series**, 19
- Class Intervals**, 36
- Collection of Data**, 7
- Correlation**, 57 *et seq.*
 - absence of, 66
 - coefficient of, 58
 - interpretation of, 61 *et seq.*
 - definition of, 57
 - degrees of, 66

Correlation—*continued*

- direct, 57, 62
 - inverse, 57, 64
 - measurement of, 58
 - perfect direct, 62
 - perfect inverse, 64
 - scatter diagrams, 69
- Cumulative Frequency Distribution**, 22
- Cyclical Movements**, 74

D

- Data**
 - arrangement of, 76
 - collection of, 7
- Deviation** (*and see* **Dispersion**)
 - average, 47
 - quartile, 44
 - standard, 50
- Dispersion**, 41 *et seq.*
 - average deviation, 47
 - coefficients of, 53
 - deciles, 44
 - definition of, 41
 - measures of, 43
 - nature of, 42
 - percentiles, 44
 - quartile coefficient of, 47
 - quartile deviation, 44
 - quintiles, 44
 - range, 43
 - semi inter-quartile range, 44
 - skewness, 55
 - standard deviation, 50

E

- Error**, 137 *et seq.*
 - classes of, 138
 - estimation of, 139
- External Statistics**, 6

F

- Frequency**
 - meaning of, 18
 - distribution, 18 *et seq.*

G

- Geometric Mean**, 38 (*and see* **Average**)
- Graphic Presentation**, 99 *et seq.*
 - rules for, 100
 - types of, 105
- Guessed Mean**, 27

H

- Harmonic Mean**, 40
- Histograms**, 114
- Historigrams**, 114

- I**
- Index Numbers**, 141 *et seq.*
Internal Statistics, 3
Interpolation
 to find median, 33
 to find quartile deviation, 46
- L**
- Line Graphs**, 105
 rulings of, 114
Logarithmic Charts, 108
- M**
- Market Research**, 10 *et seq.*
Mean (*see Average*)
Median, 31 *et seq.*
Mode, 35 *et seq.*
- P**
- Percentage Frequency of Distribution**
 23
Pictograms, 130
- Q**
- Quartile Deviation**, 44
- R**
- Rate of Change Charts**, 111
 construction, 112 *et seq.*
Ratio Graph, 109 *et seq.*
- S**
- Sampling**, 134 *et seq.*
Score Sheet, 21
Seasonal Variations
 measurement, 90
 monthly average method, 90
 moving average method, 92
 nature of, 74
- Semi-Logarithmic Charts**, 108
- Series**, 418
Silhouette Charts, 114
Spatial Distribution, 19
Standard Deviation, 50
Statistical Series, 18
 category series, 19
 frequency distribution, 18, 20
 spatial distribution, 19
 time series, 19
- Statistical Table**
 nature of, 16
 parts of, 17
 specimen, 18
- Stock Exchange List**, 7-9
- T**
- Tally Sheet**, 21
Time Series, 19, 71 *et seq.*
 classification of, 73
 charts, construction, 112
- Trend**, 71 *et seq.*
 analysis of time series, 72
 classification of time series, 73
 cyclical movements, 74
 formula, 84
 irregular variations, 77
 measurement of, 77 *et seq.*
 freehand curve-fitting method, 78
 least squares method, 83
 moving averages method, 81
 semi-average method, 78
 seasonal variations, 74
 measurement of, 90 *et seq.*
 secular, 73
 elimination of, 89
- U**
- Unbiased Error**, 138
- W**
- Weighted Arithmetic Averages**, 28

