Exercises and Worked Examples in Statistics

SECOND EDITION

G E SKIPWORTH

Exercises and Worked Examples in Statistics

G. E. Skipworth, B TECH FSS AFIMA
Senior Lecturer in Statistics at Teesside Polytechnic

Second Edition

HEINEMANN EDUCATIONAL BOOKS
LONDON

Heinemann Educational Books Ltd
22 Bedford Square, London WC1B 3HH

LONDON EDINBURGH MELBOURNE AUCKLAND
HONG KONG SINGAPORE KUALA LUMPUR NEW DELHI
IBADAN NAIROBI JOHANNESBURG
EXETER (NH) KINGSTON PORT OF SPAIN

ISBN 0 435 53791 1

First published 1971
Reprinted 1973, 1976, 1977
Second edition 1980

Printed and bound in Great Britain by
Biddles, Ltd, Guildford, Surrey

Preface to the Second Edition

This second edition is based very firmly on the first edition, but as might be expected it has been necessary to make some changes. The first four chapters contain very few alterations, although some examples have been replaced by more suitable ones. Chapter 5 is a new chapter containing examples on Time Series, Index Numbers, and Standardized Rates, topics which were omitted from the first edition but which it is now felt should be included. Chapters 5 and 6 of the first edition have been amalgamated to form a single chapter.

Perhaps the most significant change is the inclusion at the beginning of each chapter of a section (Section A) of short answer questions which are intended to test the understanding of concepts and ideas without much dependence upon numerical work. They should prove useful to students intending to sit GCE examinations containing questions of a similar type.

The author is grateful for the suggestions received over the years, and further suggestions for improvements and notification of errors are still very welcome. The author also wishes to thank Mrs M. Warnes for typing the new material used in this edition.

1980 G. E. S.

Preface to the First Edition

This book contains exercises in statistics, suitable for students taking the General Certificate of Education, students studying for Higher National Certificates and Diplomas, and undergraduates requiring a knowledge of statistical techniques without rigorous proofs.

The exercises are designed to give the student practice in statistical calculations and at the same time test the student's knowledge of fundamental principles without asking for proofs. Topics covered include probability, regression, and other branches of statistics.

It is intended that the student should use these exercises to supplement the standard textbooks; the exercises are taken from several disciplines in the hope that a student using a standard textbook, not specifically directed towards his or her discipline, will find suitable exercises in this book.

The final chapter of the book contains a number of worked examples which are intended to help a student when learning to carry out basic statistical calculations, and to this same end standard statistical tables are included. These are to be found in the section following the answers section.

The author is indebted to Dr P. W. Lewis and John Peel for the data they kindly provided, to Philip Wade for performing numerous calculations, and to Miss Irene Lindup for typing the manuscript. Suggestions for improvements and notification of errors would be gratefully received.

1971 G. E. S.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the following examination boards for their kind permission to use examples from their past examination papers:

University of Cambridge Local Examinations Syndicate [C]
Associated Examining Board for the General Certificate of Education [A.E.B.]
Oxford and Cambridge Schools Examination Board Cambridge [OC]

The author also wishes to acknowledge the permission given to him by the above-mentioned Examination Boards allowing him to convert certain of their questions to metric units and thus conform to the movement towards metrication already evident in this country.

Contents

1: *Descriptive Statistics*

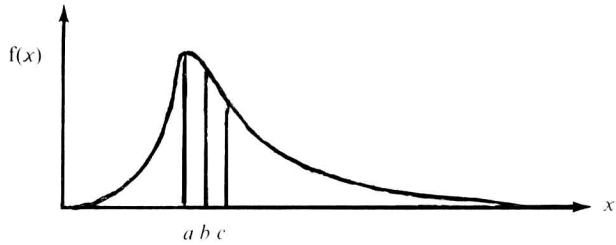
Section A

1: State briefly three purposes of pictorial representation.

2: Calculate the variance of the following values:

1 5 8 7 4

3:



For the above distribution, does *a* correspond to the mean, median or mode? What do *b* and *c* correspond to?

4: State the principle involved when drawing a histogram.

5: State one advantage and one disadvantage of using the mode as the average for a set of data.

6: Find the median of the following set of observations:

19 4 6 3 8 12 14 9

7: Write down the class boundaries, mid-point and class width for each of the following classes:

(a) 10–19,
(b) 3·5–5·5,
(c) (–6)–(–3).

8: The heights of a group of seedlings are measured to the nearest centimetre, and the results given below:

Height:	2–3	4–5	6–7	8–10	11–15
Frequency:	4	17	27	12	3

Find the minimum possible value of the range of this distribution.

9: The observations 3, x, y, 13 are in ascending order. If the median is 6 and y is twice as large as x, find the values of x and y.

10: In a year a builder uses a total of 180 tons of sand, cement and gravel, and he knows that he uses twice as much sand as gravel. He represents this information on a pie chart and the sector representing cement has an angle of 60°. How much sand does he use?

11: Sketch a frequency curve to represent data which has a moderately negatively skew distribution. Give an example of data which has such a distribution.

12: Define the semi-interquartile range and explain under what circumstances this could be useful as a measure of dispersion.

13: The average weight of a group of 8 people is 64 kg. If another four people are added to the group and their individual weights are 61, 73, 68, and 78 kg respectively, what is the average weight of all twelve people?

14: The numbers $x_1, x_2, \ldots, x_n$ have a mean of 4 and a variance of 9. What is the mean and standard deviation of the numbers:

(a) $2x_1 - 1, 2x_2 - 1, \ldots, 2x_n - 1$,

(b) $ax_1 + b, ax_2 + b, \ldots, ax_n + b$.

15: The following table shows the marks gained by a group of candidates in an examination:

Mark:	0–19	20–29	30–39	40–44	45–49	50–59	60–69	70–99
Frequency:	8	15	32	50	65	42	25	6

If this data is to be represented by a histogram calculate the heights of the histogram bars corresponding to each of the classes.

16: Calculate the mean deviation of the following numbers:

86 71 37 38 9 68 16 67.

17: The following distribution of marks was obtained with an examination paper:

Marks:	< 20	20–29	30–39	40–44	45–49	50–59	60–69	70–99
Frequency:	2	8	14	26	28	47	13	10

If this data was to be represented by a histogram, two problems would arise resulting from the division of the marks into the above

classes. What are these two problems and what would be their effect on the drawing of the histogram? One of these problems would also cause difficulty when calculating the standard deviation. Explain which problem would cause difficulty this time and what alternative measure of dispersion you would calculate in order to overcome this difficulty.

18: Calculate the Geometric Mean of the following numbers:

4 9 12 17

19: What are the advantages and disadvantages of using a pie chart to represent data.

20: Two girls, Mary and Betty sat examinations in Pure Mathematics and Applied Mathematics, and their marks were as follows:

	Pure	*Applied*
Mary	80	20
Betty	46	44

Calculate the Arithmetic Mean and the Geometric Mean of the marks for both Mary and Betty. Comment on the effect of using the Geometric Mean as an alternative to the Arithmetic Mean when assessing the mathematical ability of the two girls.

Section B

1: At 1st January, 1962 a firm employed 90 staff of whom 79 were men. During the year 17 staff left and 13 of these were men. The total recruitment during the year was 13 of whom 3 were women.

During 1963 wastage declined by 3 amongst men compared with 1962 and no women left. 6 more men, but 2 fewer women were recruited than in the previous year. The total number employed at 1st January, 1964 amounted to 93.

Arrange the above information in concise tabular form showing all relevant totals and sub-totals. [A.E.B.]

2: (a) In what circumstances is it advisable to use
(i) a histogram, (ii) a bar chart, (iii) a pie chart,
to represent statistical data?

(b) Select the most appropriate of the above methods to present the following information diagrammatically.

Production of Motor-cycles in 1961

Type of Motor-cycle	*Number (thousands)*
Mopeds (engines not exceeding 50 c.c.)	24·4
Motor-scooters	30·1
Other motor-cycles (including three-wheeled vehicles other than tractors, invalid carriages, and electrically powered vehicles)	95·3

[A.E.B.]

3: In Bournville the age distributions are as follows:
(totals in 100's (0–4 means 'under 5', 5–14 means '5 and under 15', and so on)).

Age Group	0–4	5–14	15–24	25–34	35–49
Males	71	136	108	85	125
Females	67	128	104	95	135

Age Group	50–59	60–69	70–79	80 and over
Males	70	48	29	9
Females	79	59	37	13

Display these two sets of figures (M and F) on the same histogram using two different colours.
Make two deductions or comments. [C]

4: (a) List the advantages and disadvantages of presenting information by means of bar charts and pie charts.
(b) The following is a list of the subjects taught in a school and the time devoted to each subject per week. Present the information in the form of a pie chart of 5 centimetre radius.

Subject	*Time (hours)*
English	3
French	3
Mathematics	4
Physics	$2\frac{1}{2}$
Chemistry	$2\frac{1}{2}$

[A.E.B.]

5: (a) State briefly what kind of data are best illustrated by the following diagrams:

(i) pie chart,
(ii) bar chart,
(iii) histogram.

(b) The amount of vegetables imported into the United Kingdom in the years 1953, 1957, and 1961 is given in the following table:

		Thousand tonnes		
		1953	1957	1961
Vegetables	Potatoes	121	254	261
	Onions	199	220	222
	Tomatoes	185	204	158
	Other	55	67	88
	Totals	560	745	729

Use the above table:

(i) to represent the proportions of vegetables imported in 1961 in a pie chart of radius 5 centimetre,
(ii) to calculate the radii of the circles which would effect a true comparison with the circle already drawn for the fresh vegetables imported in 1953 and 1957 if the areas of the circles are to be proportionate to the total amount of fresh vegetables imported in a year. [A.E.B.]

6: Draw diagrams to represent the following numerical data. Use a different type of diagram in each case and give brief reasons for the representation chosen.

(a) Analysis of expenditure per £100 of a motor company:.

	£
Wages	25
Materials	22
Taxation	36
Profits	13
Administration	4
	100

(b) Annual production of cars by the company:

Year:	1959	1960	1961	1962	1963
No. produced:	10 300	12 200	11 800	14 500	17 600

[C]

7: During a cricket season an all-rounder returned the following batting and bowling figures:

Batting: 25 30 6 13 0 52 10 40 8 21

Bowling: 1 for 15; 3 for 8; 7 for 42; 0 for 15; 1 for 12; 4 for 10; 6 for 72; 1 for 9; 0 for 3; 4 for 41.

What were his final averages?

8: During a course of dental treatment a patient recorded the length of time he was kept in the waiting room before the start of each treatment. The times, recorded in minutes, were as follows:

25 2 10 12 11 15
31 5 9 9 18 4

If the average treatment time was 14 minutes and all appointments were booked for this patient at 3 p.m. what was the average time at which he left the surgery?

If his bus home left at 4 p.m. and the treatment took 20 minutes what was the average time he had to wait for his bus?

9: Calculate the means of the first n integers, the first m cubes, and the first p squares.

10: In one of A. E. Housman's poems, the number of letters in each word was counted. The results were as follows:

Word Length	*Frequency*
1	4
2	13
3	20
4	13
5	8
6	9
7	1
8	3

Calculate the average word length of this poem.

11: (a) A solid rectangular block has a square base whose edge is measured to be of length 6·5 cm while the height of the block is measured as 8·0 cm. What do these measurements imply with regard to the limits of:

(i) the perimeter of the base,

(ii) the volume of the block?

(b) Briefly outline the guiding principles you would employ in tabulating statistical material.

Design a blank table to show the numbers of male and female representatives of the two major political parties elected at the General Elections held in 1965, 1966, and 1970. [A.E.B.]

12: State under what circumstances the geometric mean is used instead of the arithmetic mean.

Suppose that the turnover for a particular year has increased during 6 years by the following amounts.

Year:	1963	1964	1965	1966	1967	1968	1969
Turnover (£):	2 000	2 500	5 000	7 500	10 500	15 000	20 000

Calculate the average rate of turnover growth for this firm.

13: The lives of 50 electric lamps in hours, to the nearest hour, are given in the following table:

695	716	730	689	689	700
726	662	681	724	676	732
676	697	710	694	715	738
696	696	682	699	714	707
697	710	660	703	717	692
698	684	695	682	721	708
722	692	717	656	696	701
699	705	680	702	690	663
694	671				

Form a frequency distribution by grouping these values with a class interval of 10 hours, and draw a histogram of the distribution.

By drawing a cumulative frequency curve, or otherwise, estimate the median, mode, and 8th decile of the distribution. [OC]

14: Sales of footwear by a store in 52 consecutive weeks were:

37	60	67	63	69	54	68
60	62	83	66	70	68	61
74	94	87	66	69	66	98
62	78	90	47	70	68	98
40	73	93	51	70	71	
56	56	58	57	47	76	
59	64	46	53	54	67	
80	79	77	77	49	73	

Reduce these results to a frequency distribution with intervals centred at 40, 50, 60, etc., articles.

Represent the results as a histogram drawn on a scale of 2 cm to each interval.

On a separate diagram plot the frequencies up to the ends of each interval and sketch in a cumulative frequency curve. From this curve read off the median and quartiles of the distribution. [OC]

15: The table gives the weights in grammes of 48 golf balls measured to the nearest centigramme:

45·61	45·08	45·00	45·26	45·42	45·38
45·38	45·50	45·45	45·05	45·38	45·47
45·29	45·29	45·42	44·79	45·20	44·76
45·34	45·66	45·34	45·29	45·48	45·31
44·92	45·46	45·41	44·98	45·69	45·37
45·21	45·55	45·44	45·68	45·02	45·53
45·04	45·02	45·37	45·28	45·34	45·47
45·15	45·52	45·38	44·95	45·57	45·14

Divide the range of variation into equal intervals of which the first is 44·70 to 44·799 g, calculate the frequency in each interval, and draw a histogram of the weights.

Draw a cumulative frequency curve to fit the population of which these weights are a sample and from this curve read off the median weight and the values of the quartiles. [OC]

16: The speeds in kilometres per hour of vehicles passing along a road were measured and the results given in the following table:

Centre of Interval:	25	30	35	40	45	50	55	60	*Total*
Frequency:	5	25	72	122	180	256	118	22	800

Illustrate these results by drawing

(i) a histogram,

(ii) a cumulative frequency curve.

From the graphs estimate the median, the semi-interquartile range, and the percentage of vehicles travelling at speeds greater than 55 km.p.h. [OC]

17: An analysis of 500 sentences in a certain book gave the following results:

No. of Words:	1–15	16–30	31–45	46–60	61–75
Frequency:	36	189	212	52	11

Calculate the mean and state the range in which the upper quartile occurs. [C]

18: Percentages of a sample of the population within certain height limits are given by the table:

Height	*Percentage*
Under 153 cm	0·4
153 cm and under 158	3·0
158 cm and under 163	9·4
163 cm and under 168	15·1
168 cm and under 173	23·4
173 cm and under 178	25·3
178 cm and under 182	15·0
Over 182 cm	8·4
Total	100·0

Draw the cumulative frequency curve. Estimate the median and the interquartile range. [C]

19: The table gives the weekly wage distribution in a factory.

Wage in £:	12–	13–	14–	15–	16–	17–	18–
No. of Employees:	90	150	240	290	160	70	0

Calculate the mean weekly wage and the range in which the lower quartile occurs. [C]

20: The following distribution of marks (out of 100) was obtained with a certain examination paper:

Mark Range:	10–29	30–39	40–49	50–59	60–69	70–79	80–
Frequency:	2	8	14	26	28	10	4

Present these marks in a histogram.

Calculate the mean mark as well as you can, explaining the limitations of your calculation. Do you consider that the mean or the median is the better measure of average performance for these observations? Give reasons for your answer. [C]

21: A survey of kilometres covered by 1 000 cars gave the following results:

Kilometres	*Frequency*	*Kilometres*	*Frequency*
1–4 999	42	25 000–29 999	152
5 000–9 999	73	30 000–34 999	101
10 000–14 999	115	35 000–39 999	88
15 000–19 999	154	40 000–44 999	53
20 000–24 999	172	45 000–49 999	50

Draw a cumulative frequency curve and find the interquartile range. [C]

22: The following table shows the number of candidates C, scoring M marks for a question in an examination:

M:	0	1	2	3	4	5	6	7	8	9	10
C:	6	24	88	108	110	112	90	69	38	21	4

Calculate the mean and the median of the distribution. What feature of the distribution causes the mean to be greater than the median? [C]

23: A machine produces rods whose diameters are required to be within the tolerance limits 0·988 cm to 1·012 cm. A sample of 150 rods, measured to the nearest thousandth of a centimetre, gave the following distribution:

Diameter (cm)	0·976 to 0·981	0·982 to 0·987	0·988 to 0·993	0·994 to 0·999	1·000 to 1·005	1·006 to 1·011	1·012 to 1·017
Number of Rods	1	5	30	71	34	7	2

(i) Construct the cumulative frequency curve for the rods. (Take 2 centimetre on the graph to represent 0·01 cm and 2 centimetre to represent 20 rods.)

(ii) Calculate the percentage number of rods outside the tolerance limits. [A.E.B.]

24: Calculate a cumulative frequency table for the distribution in question 43 and draw the cumulative frequency graph. (Take 2 centimetre horizontally to represent 20 marks, and 2 centimetre vertically to represent a frequency of 200.)

From your graph estimate:

(a) the value of the median mark,
(b) the value of the pass mark if 40% of the candidates fail the examination,
(c) the semi-interquartile range. [A.E.B.]

25: Calculate:
(i) the arithmetic mean,
(ii) the median,
(iii) the mode,
of the following numbers:

1·0, 2·2, 1·1, 2·0, 1·4, 2·0, 1·4, 1·8, 1·4, 1·7. [A.E.B.]

26: State between what limits the following product lies if the numbers are rounded off to the given significant figures

$$9{\cdot}73 \times 2{\cdot}41.$$

Calculate:
(i) the arithmetic mean,
(ii) the median,
(iii) the mode,
of the following numbers:

10, 14, 22, 16, 15, 14, 15, 13, 14, 17. [A.E.B.]

27: The marks of 50 candidates in an examination for which the maximum mark was 100 are given below

62	21	4	26	7	38	32	64	12	38
45	6	33	55	62	48	49	7	9	41
21	30	31	3	25	57	48	8	18	43
72	23	5	8	37	31	31	39	65	53
4	75	17	14	61	50	51	38	36	40

Select suitable classes and draw up a frequency distribution. Draw a histogram to represent the data and write down the modal class. Comment on any significant features. [C]