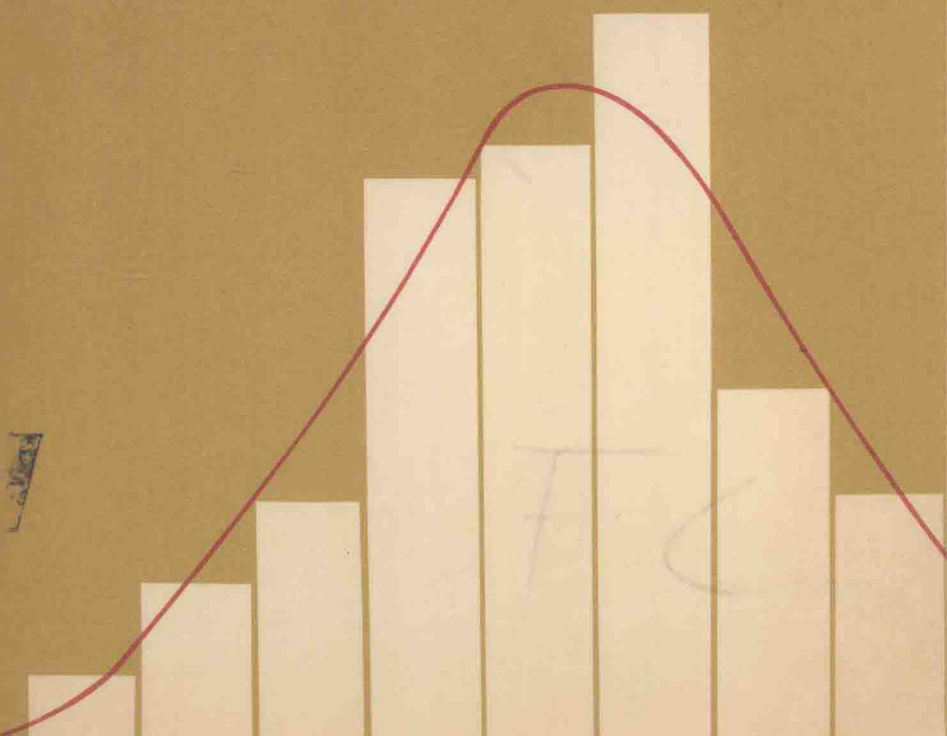
Blackwell Scientific Publications

John F Osborn

Statistical Exercises in Medical Research



STATISTICAL EXERCISES IN MEDICAL RESEARCH

JOHN F. OSBORN

BSc, PhD

Lecturer in Medical Demography,

London School of Hygiene and

Tropical Medicine

University of London

FOREWORD BY

P. ARMITAGE

Professor of Biomathematics

University of Oxford

BLACKWELL SCIENTIFIC PUBLICATIONS

OXFORD LONDON EDINBURGH MELBOURNE

Osney Mead, Oxford OX2 0EL
8 John Street, London, WC1N 2ES
9 Forrest Road, Edinburgh, EH1 2QH
P.O. Box 9, North Balwyn, Victoria 3104, Australia

First published 1979

British Library Cataloguing in Publication Data

Osborn, John F
Statistical exercises in medical research.
1. Mathematical statistics
2. Medical statistics
I. Title
519.5′02′461 QA276

ISBN 0–632–00026–0

Distributed in North America by
Halsted Press, a division of
John Wiley & Sons Inc.
605 Third Avenue, New York, NY 10016

Printed in Great Britain by
Western Printing Services Ltd,
Avonmouth
and bound by
Kemp Hall Bindery
Oxford

STATISTICAL EXERCISES IN MEDICAL RESEARCH

TO THE STUDENTS

IT HAS BEEN MY PRIVILEGE

TO TEACH

FOREWORD

Statistical methods are the techniques which have been developed for the collection, processing and interpretation of real data. Any course in statistics which avoids substantial contact with data therefore presents a very restrictive view of the subject. By studying a variety of data sets the student, whether a future statistician or a scientific research worker, is able to grasp the context in which particular methods are appropriate and to become more fluent in the use of these methods.

Textbooks on mathematical statistics usually include theoretical exercises. Those on applied statistics are less likely to include comparable practical exercises because of the space required and the task of finding sensible data. Dr Osborn has performed a valuable service in compiling the present collection. The data are taken predominantly from medical research projects, and the techniques which the reader is invited to apply cover a wide spectrum of commonly-used methods. The collection will be of great value as an adjunct to courses on medical and biological statistics, and for all those who wish to obtain more experience in the practical use of statistical methods.

P. Armitage
Professor of Biomathematics
University of Oxford

PREFACE

Teaching medical statistics to post-graduate medical students during the past decade or so, has taught me that lectures, seminars and textbooks are not sufficient to ensure that the statistical material intended to be taught, is in fact learned. At the London School of Hygiene and Tropical Medicine, each one hour lecture in medical statistics is followed by practical work lasting about one and a half hours. Despite this, it is very common for students to request further exercises to ensure that the learning objectives of the lecture are in fact realized. Thus this book of exercises is intended to help students and medical research workers who, having received instruction in the methods of medical statistics, wish to confirm their ability to apply the techniques to data. The students at the London School of Hygiene and Tropical Medicine have interests in a wide variety of medical specialities including, for example, community medicine, medical demography, epidemiology, human nutrition, occupational health, tropical medicine and so on. The exercises in this book have been selected from a similarly wide variety of subjects in order to provide a balance which, it is hoped will be of interest to all medical research workers, whatever their speciality.

Since its publication, *Statistical Methods in Medical Research* (Armitage P. (1971) Blackwell Scientific Publications, Oxford), has established itself as a standard text and reference work among both students and research workers in medicine and *Statistical Exercises in Medical Research* is intended to be a companion volume, the chapter and section titles being common to both books for ease of reference. This, however, has led to some anomalies in the numbering of some sections in *Statistical Exercises in Medical Research*. For example, Chapter 1 starts at section 1.2, since in *Statistical Methods in Medical Research* section 1.1 is devoted to a general discussion of the application of statistical methods to medical problems, for which it is difficult to envisage appropriate statistical exercises for the reader. Section 1.2 in both books is headed Diagrams.

JOHN F. OSBORN

ACKNOWLEDGEMENTS

In any book such as this, which is to a large extent, the result of the author's teaching experience, any complete list of acknowledgements would be excessively long. I am very grateful to all my colleagues, friends and students who have made contributions to this book. However, several people have given very specific help and, in particular, I must thank Professor Peter Armitage who first suggested to me that a book of statistical exercises would be useful to medical students and medical research workers, and who has helped and encouraged me throughout its preparation. Some of the exercises have been used at the London School of Hygiene and Tropical Medicine in the courses of medical statistics for many years, and I am grateful to Professor Armitage, Professor William Brass, Dr Paul Fine, Dr Ronald H. Gray and other colleagues who originally abstracted this data from the literature. For assistance with the checking of answers and advice on computing I thank Miss Patricia E. Doyle and Ms Basia Zaba. The typescript was read by Professor Armitage, Professor Brass, Professor Michael R.H. Healy, my wife, Mrs Erica M. Osborn and my father Mr Frederick J. Osborn, all of whom made suggestions which improved the exercises.

I am also grateful to Miss Sandra A. Etheridge whose expertise has enabled the manuscript to be typed quickly and accurately. Miss Dorothy E. Boyes and Miss Sue M. Evans gave additional secretarial help.

Finally, I must thank all those workers in medical research whose efforts have produced the data for these exercises. References to the origin of data have been given where these have been traced, but some of the material, particularly data which have been tailored for use as teaching material in the London School of Hygiene and Tropical Medicine for many years, is given without reference.

If, despite the valuable assistance I have received, there are residual errors, these are entirely my responsibility, and I would be grateful to have them pointed out to me.

JOHN F. OSBORN

CONTENTS*

* Chapter and section numbers and titles correspond to those in Armitage P. (1971) *Statistical Methods in Medical Research*, Blackwell Scientific Publications, Oxford.

CHAPTER 1

THE SCOPE OF STATISTICS

1.2 DIAGRAMS

1.2.1 The data below show the numbers of neonatal and post-neonatal deaths recorded during each month of 1951 in England and Wales. (A neonatal death is the death of an infant within 28 days of birth; a post-neonatal death is the death of an infant aged between 29 days and one year.) Calculate the average number of neonatal and post-neonatal deaths per day, for each month and plot these to indicate the differences in seasonal trend.

Month	Neonatal deaths	Post-neonatal deaths
January	1224	1080
February	1099	872
March	1271	923
April	1141	705
May	1137	553
June	1077	410
July	959	408
August	952	381
September	907	390
October	982	444
November	973	493
December	1036	699
Total	12,758	7358

1.2.2 The crude birth rate (CBR) and the crude death rate (CDR) for England and Wales from 1926 are shown below. The difference between these two rates is called the crude rate of natural increase (CRNI). Plot graphs of these three rates to show their trends since 1926.

Rates per 1,000 population				Rates per 1,000 population			
Year	CBR	CDR	CRNI	Year	CBR	CDR	CRNI
1926	17·8	11·6	6·2	1952	15·3	11·3	4·0
1927	16·7	12·3	4·4	1953	15·5	11·4	4·1
1928	16·7	11·7	5·0	1954	15·2	11·3	3·9
1929	16·3	13·4	2·9	1955	15·0	11·7	3·3
1930	16·3	11·4	4·9	1956	15·7	11·7	4·0
1931	15·8	12·3	3·5	1957	16·1	11·5	4·6
1932	15·3	12·0	3·3	1958	16·4	11·7	4·7
1933	14·4	12·3	2·1	1959	16·5	11·6	4·9
1934	14·8	11·8	3·0	1960	17·1	11·5	5·6
1935	14·7	11·7	3·0	1961	17·6	11·9	5·7
1936	14·8	12·1	2·7	1962	18·0	11·9	6·1
1937	14·9	12·4	2·5	1963	18·2	12·2	6·0
1938	15·1	11·6	3·5	1964	18·6	11·3	7·3
1939	14·8	12·1	2·7	1965	18·1	11·5	6·6
1940	14·1	14·4	−0·3	1966	17·8	11·7	6·1
1941	13·9	13·5	0·4	1967	17·3	11·2	6·1
1942	15·6	12·3	3·3	1968	16·9	11·9	5·0
1943	16·2	13·0	3·2	1969	16·4	11·9	4·5
1944	17·7	12·7	5·0	1970	16·1	11·7	4·4
1945	15·9	12·6	3·3	1971	16·0	11·6	4·4
1946	19·2	12·0	7·2	1972	14·8	12·0	2·8
1947	20·5	12·3	8·2	1973	13·7	11·8	1·9
1948	17·8	11·0	6·8	1974	13·0	11·8	1·2
1949	16·7	11·8	4·9	1975	12·2	11·7	0·5
1950	15·8	11·6	4·2	1976	11·9	12·0	−0·1
1951	15·5	12·5	3·0				

1.2.3 The data below show the percentage of the population of England and Wales who are of pensionable age (i.e. males aged 65+ and females 60+ years of age).

(a) Draw a graph to illustrate the trend in the percentage of persons of pensionable age during the period 1901–1971.

(b) How might this trend affect the trend in the crude death rate during this period? (See for example exercise 1.2.2.)

Year	1901	1911	1921	1931	1941	1951	1961	1971
%	6·2	6·8	7.9	9.6	11·8	13·6	14.6	16·0

1.4 SUMMARIZING NUMERICAL DATA

1.4.1 The recorded birth weights of 18,645 singleton live and stillbirths occurring in South-West England in 1965 are given below. There is some evidence of digit preference particularly at zero and 8 ounce points. Using one pound weight intervals construct a frequency distribution and a relative frequency distribution. Draw a histogram to illustrate the data.

Distribution of Recorded Birth Weight, South-West of England, 1965
Number of Singleton, Live and Stillbirths

Pounds	Ounces															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2
1	6	1	1	1	3	0	2	2	3	1	3	4	8	2	2	1
2	18	4	2	2	6	2	4	2	10	4	4	2	8	7	4	3
3	14	6	8	5	9	6	8	9	14	2	6	6	7	5	14	7
4	22	14	16	19	16	14	15	19	47	17	23	15	39	30	26	32
5	66	37	42	46	60	41	67	59	106	78	98	68	135	92	106	81
6	323	101	183	157	337	160	205	172	504	215	299	222	496	256	315	228
7	914	225	390	286	697	311	417	291	817	289	369	279	626	246	330	236
8	920	195	292	220	508	200	230	166	485	147	198	110	288	122	146	78
9	395	83	118	72	142	53	69	45	145	35	42	22	91	18	25	10
10	88	12	26	9	23	11	6	4	18	8	7	2	16	4	2	4
11	17	1	3	2	3	1	0	2	2	0	4	1	2	0	1	0
12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Total, including 51 with unknown birth weight: 18,696

Source: Pethybridge R.J., Ashford J.R. & Fryer J.G. (1974) *Brit. J. prev. soc. Med.* **28,** 10–18.

1.4.2 The distributions in the first table below show the pre-operational percentage haemoglobin values of a sample of the population of a village where there has been a malaria eradication programme (MEP).

The results in a sample obtained after MEP are given in the second table.

Construct similar distributions for these post-operational percentage haemoglobin values. Compare the relative frequency distribution observed before MEP with that obtained after MEP.

Haemoglobin %	30–39	40–49	50–59	60–69	70–79	80–89	90–99	100–109	Total
Frequency	2	7	14	10	8	2	2	0	45
% of total	4·4	15·6	31·1	22·2	17·8	4·4	4·4	0	99·9

43	63	63	75	95	75	80	48	62	71	76	90
51	61	74	103	93	82	74	65	63	53	64	67
80	77	60	69	73	76	91	55	65	69	84	78
50	68	72	89	75	57	66	79	85	70	59	71
87	67	72	52	35	67	99	81	97	74	61	72

1.4.3 Detailed studies on filariasis in Fiji required that a census be made of the population. Data for 8 villages on the islands of Taveuni and Koro are given below. (*Report to the Director of Medical Services, Fiji*, 1968–1969.)

Draw a histogram to show the percentage distribution of the population of males by age.

Draw a cumulative relative frequency distribution. From the graph of the cumulative distribution determine the age which divides the population 50–50, i.e. at what age, say x years, are 50% of the males younger than x and 50% older than x?

Age last birthday	No. males	% of total males
0–4	154	18·6
5–9	135	16·3
10–14	107	12·9
15–19	72	8·7
20–29	112	13·5
30–39	97	11·7
40–49	67	8·1
50–59	47	5·7
60–79	39	4·7
Total	830	100·2

1.4.4 Using 2 mm grouping intervals, construct a frequency distribution and relative frequency distribution of the skinfold thicknesses given in the table below. The measurements given are skinfold thicknesses in millimetres at the triceps mid-point for 121 male subjects.

11·4	15·3	9·1	18·4	10·9	4·7	9·6	20·6	10·4	20·5	22·4
14·3	11·7	11·4	12·7	18·2	15·1	14·6	25·3	11·5	13·2	7·9
12·6	13·9	16·8	11·4	27·3	16·3	13·9	13·2	11·9	20·0	13·2
9·4	18·9	10·7	14·8	17·8	10·8	16·0	15·7	17·7	13·5	11·5
11·1	9·6	15·1	13·6	13·6	8·6	6·9	19·1	18·7	10·1	16·0
20·4	7·9	16·6	18·5	16·2	17·4	18·8	12·6	22·0	9·6	11·1
15·7	23·7	13·3	4·9	8·3	20·1	15·5	23·1	10·2	10·7	15·8
17·6	21·3	16·2	14·9	9·9	9·1	9·9	9·8	8·6	11·8	9·3
14·8	17·3	9·5	13·6	12·4	9·5	14·3	25·7	12·9	22·7	12·1
10·7	16·8	11·3	11·3	11·4	5·9	10·7	14·6	19·8	25·5	7·7
18·4	7·9	7·6	23·3	9·6	8·4	10·4	8·1	12·5	9·0	30·1

Source: Colley J.R.T., personal communication; *see also* Ruiz L., Colley J.R.T. & Hamilton P.J.S. (1971) *Brit. J. prev. soc. Med.* **25,** 165–167.

1.4.5 The sprayable surface area (in square feet) of the 250 houses of an African village are shown below. Using class widths of 10 square feet construct a frequency distribution of the sprayable surface areas.

No.	Area	No.	Area	No.	Area	No.	Area	No.	Area	No.	Area	No.	Area
1	216	41	328	81	185	121	236	161	254	201	286	241	173
2	231	42	262	82	233	122	219	162	238	202	192	242	208
3	194	43	282	83	214	123	255	163	274	203	300	243	283
4	188	44	221	84	237	124	245	164	265	204	276	244	253
5	326	45	243	85	271	125	257	165	234	205	291	245	227
6	214	46	241	86	263	126	191	166	224	206	191	246	242
7	264	47	203	87	261	127	284	167	212	207	291	247	231
8	255	48	168	88	307	128	211	168	249	208	290	248	250
9	242	49	258	89	215	129	293	169	254	209	211	249	229
10	305	50	234	90	261	130	184	170	234	210	280	250	287
11	311	51	314	91	256	131	219	171	217	211	258		
12	300	52	259	92	224	132	272	172	247	212	200		
13	254	53	301	93	259	133	251	173	206	213	234		
14	193	54	219	94	221	134	272	174	283	214	263		
15	245	55	282	95	239	135	254	175	264	215	236		
16	165	56	303	96	230	136	271	176	332	216	245		
17	201	57	207	97	257	137	222	177	219	217	254		
18	241	58	283	98	237	138	266	178	262	218	249		
19	220	59	204	99	238	139	281	179	319	219	258		
20	240	60	264	100	257	140	227	180	338	220	207		
21	281	61	281	101	296	141	259	181	309	221	225		
22	261	62	270	102	266	142	235	182	318	222	252		
23	320	63	196	103	285	143	223	183	327	223	281		
24	240	64	204	104	315	144	243	184	298	224	253		
25	259	65	233	105	237	145	255	185	196	225	210		
26	292	66	239	106	234	146	268	186	262	226	228		
27	184	67	255	107	181	147	282	187	242	227	230		
28	320	68	325	108	193	148	253	188	261	228	261		
29	269	69	289	109	256	149	226	189	258	229	194		
30	277	70	263	110	247	150	248	190	305	230	304		
31	285	71	251	111	269	151	303	191	285	231	227		
32	318	72	246	112	218	152	301	192	256	232	243		
33	258	73	222	113	329	153	254	193	228	233	271		
34	182	74	205	114	227	154	244	194	209	234	326		
35	199	75	170	115	284	155	228	195	244	235	278		
36	294	76	209	116	255	156	271	196	251	236	272		
37	286	77	310	117	250	157	269	197	232	237	267		
38	203	78	272	118	269	158	316	198	171	238	285		
39	266	79	288	119	252	159	261	199	255	239	298		
40	219	80	214	120	247	160	287	200	213	240	245		