Institute of Medical Laboratory Sciences Monographs

Medical Laboratory Statistics

Paul W Strike

Medical Laboratory Statistics

Paul W. Strike FIMLS, Dip. Biom., FSS

Biochemist,
R.A.F. Institute of Pathology
and Tropical Medicine

WRIGHT · PSG
Bristol London Boston
1981

Published by:
John Wright & Sons Ltd., 42–44 Triangle West, Bristol BS8 1EX, England.

John Wright PSG Inc., 545 Great Road, Littleton, Massachusetts 01460, U.S.A.

British Library Cataloguing in Publication Data

Strike, Paul W.
Medical Laboratory statistics
1. Mathematical statistics
2. Medical statistics
I. Title
519.5'02461 QA276

ISBN 0 7236 0582 3

Printed in Great Britain by
John Wright & Sons Ltd., at The Stonebridge Press, Bristol BS4 5NU

Foreword

With the rapid advancement of medical laboratory sciences, it is often difficult for the laboratory worker to keep abreast of current knowledge. Major text books cover a wide field of a given major discipline, but obviously they cannot cover every facet of a subject.

The Institute of Medical Laboratory Sciences Monograph Series is written by experts in specific areas, to expand on a subject and to give it the depth and breadth not generally possible in a major work. These books will be useful to those working in the medical laboratory sciences field, for use on the bench, as an examination reference book and for keeping up with current thinking.

I am delighted to have been offered the General Editorship of this series and hope it will be the success it well deserves to be. It is hoped that the Monograph series will be extant for a long period of time and that it will build up to be a useful and welcome addition to your library.

F. J. Baker
London 1981

Preface

The content of this slim volume has been to some extent determined by first-hand experience of the midnight blood barbiturate. What seemed important then and in the small hours of the night is reflected in what follows. The selection is personal, a blend of workshop manual and contemplative essay, with some sixty illustrations for the numerically faint of heart. The level of mathematics demanded never passes beyond simple arithmetic, although there is a good deal of it at times. Much of this can be left to a programmable calculator or computer. It is in the framework of ideas and assumptions within which the trivial arithmetic is performed that the real difficulties lie.

Read this book from cover to cover. Do not dip into it until you have read it at least once. The ideas are developed sequentially, and for the most part pictorially. Difficult areas are continually revisited from different standpoints. The subject of regression in particular is dealt with in a detailed manner, given its central role in the measurement process. Given the constraints upon size and the particular emphasis of the monograph on laboratory measurement, the chi-squared tests have been excluded. A clear and inexpensive introduction can be found in Sprent P. (1977) *Statistics in Action*, Middlesex, Harmondsworth; Penguin.

The book covers most of the requirements of the Institute of Medical Laboratory Sciences Fellowship examinations, and should prove useful to candidates for the Association of Clinical Biochemists Mastership in Clinical Biochemistry, the Royal College of Pathologists Membership examinations and to anyone concerned with the medical laboratory.

It is a pleasure to acknowledge the help of Wing Commander S. A. Cullen of the Royal Air Force pathology branch who had the questionable pleasure of reading everything, much of which never made it to the final draft, thanks to his patient criticism. Professor M. J. R. Healy, of the London School of Hygiene and Tropical Medicine, read an early draft with sharp eye and good humour. My wife Jennifer typed a difficult and ever-changing manuscript, removing a good deal of nonsense on the way.

P.W.S.

Contents

1 Introduction

One thing the scientific discipline of statistics does not lack is a multiplicity of definitions for itself. Barnett (1975) advanced the following provisional definition which seems well suited for this book. Statistics is 'the study of how information should be employed to reflect on, and give guidance for action in, a practical situation involving uncertainty.' It allows us to summarize, manipulate and base inference upon sample data that have an element of uncertainty either as an inherent property or as a consequence of the manner in which they were obtained.

The word 'statistics' has its origins in the collection of information for the State. The military adventures of the 17th century demanded finance and the one main source of this finance, taxation, required a knowledge of taxable assets. 'State information' rapidly infiltrated many of the affairs of government. The 17th century also saw the birth of a new branch of mathematics, probability theory, with its roots in the gaming houses of Europe. In less than a hundred years the approximation of State-istics by probability models was a reality. The development of Statistics owes much to the scientists of the day who recognized its potential importance in their own fields. The astronomer Karl Friedrich Gauss (1777–1855) developed the method of least-squares, a fundamental contribution to the analysis of errors of observation. His essays (recorded in Latin) are a salutary reminder that microprocessors have a long way to go no matter how good a game of chess they may play.

In the past hundred years the science has developed to a considerable level of complexity and subtlety thanks to the contributions of scientists and mathematicians alike. Its reputation has not always matched its substance. In the field of government and economic forecasting in particular it has been, and is seen to be, fallible. Economic situations are complex and for entire nations, the subject of many complex material, social and political interactions. Even after the most exacting statistical analyses, next year remains as much of an unknown quantity as this year was twelve months ago.

A second blow to its reputation is paradoxically a reflection on its widespread value. With applications in so many diverse areas—the sciences, economics and industry—it is hardly surprising that a good number of people, armed with any one of the legion recipe books available, attempt their own analyses. The medical journals bear eloquent witness to the quality of many of these enterprises.

Coming to study statistics from a background in the medical laboratory sciences, one is quickly impressed by the contrast between the simplicity of the calculations and the complexity of the reasoning. If you can add two and two you will soon find your way through the arithmetic of a *t*-test. The logical principles underlying the rejection of one hypothesis in favour of another, on the basis of that *t*-test, are by no means as easy to grasp; indeed, experts still find the subject an ever ready source of argument and controversy.

Numbers are tricky customers at the best of times. Completely non-existent problems can be conjured out of the most innocent situations simply by adopting the wrong point of view. The following paradox, borrowed from Northrop (1975) is a trivial example of the point. Three men dining together at a restaurant receive a bill for £30. Each man gives the waiter £10 who returns to the cash desk. Here he is told that a mistake has been made over the bill, which should have been £25. He is given £5 to return to the men. Returning to their table, the waiter reasons that £5 is an awkward amount to divide between three men and they of course would be only too pleased to get anything back at all, so, he pockets £2 and gives each of the men £1. Now, each man has paid £9 for his meal. Three times nine is twenty-seven. The waiter has two pounds in pocket. Twenty-seven plus two is twenty-nine. The men originally handed over thirty pounds. Where is the missing pound?

The explanation is obvious given a moment's reflection. Lose this problem in a mass of superfluous facts and figures and the necessary re-orientation may be a lot more difficult to achieve. The point being made is lent a rather chilling reality by the following story, first related by Dr Richard Asher (1954). An investigator observed that of 200 epileptic subjects studied, 24% had infantile convulsions in their first two years of life. Of 200 'normal' subjects only 2% had infantile convulsions in this period. The investigator reasoned that, given the very obvious difference between the 24% in epileptics and the 2% in normals,

convulsions in the first two years of life could be taken as strong evidence of epilepsy and the child placed immediately on anti-convulsant therapy.

It sounds plausible but something has been overlooked. Can you detect the fallacy at this point?

No mention has been made of the incidence of epilepsy in the population at large! This is approximately 1:400. In a group of 40 000 people from the general population we would expect to find approximately 100 epileptics, 24 of whom would have had infantile convulsions before they were two years old. However, among the 40 000 people we would also expect to find approximately 800 people, who did not have epilepsy, who also had infantile convulsions before they were two years old, i.e. 2% of 40 000. If we had followed this investigator's advice we would have ended up putting 800 children on long-term anti-convulsant therapy, only 24 of whom actually had epilepsy!

If this example is submerged in another two or three thousand words of text the fallacy may, once again, be far from obvious.

Modern computing devices will effortlessly perform the arithmetic; without a degree of thought they will as effortlessly lead you into trouble. The more complex the problem, the more care you will need to employ in considering both the properties of your observations (e.g. are they *really* a random sample? ... what are they a random sample of? ... are the observations contaminated by measurement errors?) and of your statistics. The computer can calculate correlation matrices for multivariate analyses in seconds, a task that in the past demanded many hours (sometimes months!) of awesomely monotonous calculation. The potential for complete dislocation from any sort of reality in these extremely complicated models is enormous. If you do not have a pretty good idea of where you are, what you are doing and where you want to go before you load the computer, the subsequent journey through a multidimensional vector space may leave you with some substantial illusions, all to ten or more decimal places!

Statistical encounters in the medical laboratory are of more fundamental importance than you might at first admit. The laboratory has much in common with manufacturing industry, processing raw materials to produce a 'product' in response to consumer demand, the product in this case being a 'result'. It is sobering to reflect for a moment upon the body of consumer legislation that would come to bear upon the laboratory if its

product were bags of crisps! As it is, the extent to which the laboratory's product, its results, satisfy its main consumer, the clinician, depends upon how well they reflect the clinical reality. Success here depends entirely upon how thoroughly the laboratory has aswered such questions as: 'How do we know we are measuring what we believe we are measuring? . . . How reproducibly do we measure whatever it is we believe we are measuring? . . . How does the inherent variability of our observations affect their diagnostic value?' These and similar questions are often quite profound. A good deal of both common-sense and statistical expertise may be called upon to provide answers that are useful and seen to be useful.

Common-sense is not within the gift of this slim volume and statistical expertise is the substance of an undergraduate discipline.

Between the two lies an awareness of what can be done and what cannot, but often is done.

Most, if not all, laboratory staff have acquired a profound suspicion of any science that ends all of its statements with $p < 0{\cdot}05$. Years of experience have taught that if there is less than a 1:20 chance that a component will fail, then it will fail in an emergency. The explanation is simple. The $p < 0{\cdot}05$ was based on an ill-defined sample space. Redefine the problem and you will get the correct answer of $p = 1{\cdot}0$, i.e. the component will fail in every emergency! This volume will teach you all this and much besides.

REFERENCES

Asher R. (1954) Straight and crooked thinking in medicine. *Br. Med. J.* **2**, 460–462.

Barnett V. (1975) *Comparative Statistical Inference*. New York, Wiley.

Northrop E. P. (1975) *Riddles in Mathematics*. Harmondsworth, Middlesex, Penguin.

2 First Steps

2.1 GETTING THE PICTURE

A variable is a property with respect to which individuals in a sample differ in some ascertainable way. Biological variables may be classified as follows:

1. *Nominal.* Also referred to as categorical, attribute or qualitative variables. These cannot be measured or ranked but are expressed qualitatively, e.g. pregnant or non-pregnant. When combined with frequencies they can be treated statistically.

2. *Ordinal.* These have an ordering or ranking significance and nothing more, e.g. the order in which ten eggs hatch; A, B, . . . J. The difference between egg A and egg B has no meaning in terms of magnitude. Sometimes nominal data may be coded in ordinal form. If the nominal variables are responses to a drug, i.e. worse, same, better, much better, they could be coded 1, 2, 3 and 4, each value representing a qualitative improvement on its predecessor.

3. *Measurement.* These have a magnitude significance, i.e. four is twice as great as two. They are divisible into:

a. Discrete measurement variables: these can assume certain fixed numerical values only with no intermediate values. In practice this implies a count;

b. Continuous measurement variables: these represent observations on a continuum (length, optical density, etc.) and theoretically an infinite number of values are realizable between any two points. In practice, limitations in the resolving power of the measuring instrument lead to the expression of continuous measurement variables in discrete form, e.g. a blood glucose concentration may be 5·078 913 . . .mmol/l: the assay method is accurate to 0·1mmol/l as a consequence of its imprecision, obliging us to express the blood glucose in discrete form, 5·1mmol/l. The underlying continuum is of considerable importance in subsequent statistical analyses.

Table 1 presents the realizations of a continuous measurement variable, the serum level sof human placental lactogen (HPL) in mg/l for 300 healthy women of child bearing age between 36 and 39 weeks of pregnancy. They were obtained over a twelve month period from five hospitals in different regions of England.

Table 1. Sample of human serum placental lactogen concentrations in 300 women between 36 and 39 weeks of pregnancy.

5·0	6·2	7·1	5·5	6·5	4·6	6·3	8·0	8·5	7·2
7·1	4·0	6·6	7·6	7·9	9·1	10·2	6·3	8·3	5·4
7·8	7·8	6·8	8·5	7·6	5·1	6·6	8·5	7·4	8·7
5·8	9·1	6·2	9·1	9·6	5·8	8·7	10·2	5·7	6·6
7·2	6·9	4·7	5·5	5·7	8·5	5·7	7·8	8·3	6·9
5·4	9·1	7·3	8·0	11·2	7·8	4·3	5·4	7·6	4·8
7·7	8·7	8·7	8·0	6·6	9·6	7·4	6·9	6·6	6·8
8·3	5·3	11·2	9·6	6·3	10·2	5·3	10·2	6·5	6·2
7·3	10·0	9·1	5·8	7·3	6·9	5·8	8·3	4·4	8·5
6·8	9·1	7·1	6·6	5·7	6·3	5·7	8·9	8·5	7·1
9·1	5·8	8·5	12·3	5·7	7·4	7·9	5·8	7·3	11·2
4·5	9·6	10·2	9·8	5·4	6·8	8·0	8·7	6·0	6·3
6·9	10·0	6·0	6·9	7·8	4·9	7·8	8·3	7·3	6·5
4·6	6·6	5·3	11·8	7·3	6·8	9·6	4·6	5·7	8·3
8·5	8·5	9·6	6·5	9·6	7·6	5·0	6·8	5·3	9·6
6·5	7·1	11·5	7·8	6·3	7·1	6·2	6·5	7·8	10·3
5·0	8·5	13·2	11·5	6·0	7·3	8·3	4·0	4·4	8·0
7·8	8·1	9·4	5·5	7·8	7·6	9·4	7·8	7·8	9·4
8·1	6·9	7·8	8·7	5·8	10·5	4·5	5·5	6·3	5·7
7·4	6·6	7·7	13·2	7·4	8·5	8·5	6·3	8·0	7·8
6·5	5·7	6·6	6·2	7·6	6·6	11·4	8·5	7·1	5·5
8·3	10·0	12·7	7·4	12·9	8·7	7·6	4·7	9·6	6·8
7·4	7·8	9·1	6·5	6·3	5·1	9·4	11·2	11·2	6·5
5·4	7·4	12·1	12·3	8·7	8·9	7·8	5·7	11·5	5·8
6·8	6·6	8·7	13·5	7·9	6·8	12·6	10·0	9·6	7·1
6·3	7·6	5·8	4·5	6·6	11·2	11·3	14·0	7·3	10·0
8·9	10·7	5·7	12·1	14·0	6·2	5·3	7·6	6·5	8·1
5·6	7·1	9·8	8·5	12·6	10·5	9·6	8·3	9·1	7·8
9·8	9·8	4·8	10·7	8·1	11·2	9·7	6·8	7·1	9·3
8·0	6·8	10·7	8·1	10·6	6·5	7·1	8·3	6·3	8·9

A casual inspection of the figures reveals very little about the structure of the data. At best we might suggest that most of the

values seem to be below ten and they are all rounded to one decimal place. Unravelling the patterns in the data (if any) and summarizing its properties in a few easily interpreted numerical measures may be all that is required. If a more complex analysis is planned these preliminaries will still be necessary in order to ascertain which techniques are appropriate, given the particular properties of the data at hand. A simple histogram may be so telling that subsequent analyses are made unnecessary.

A statistical analysis of the data in *Table 1* presupposes a question in need of an answer. It cannot be emphasized strongly enough that the formulation of this question, whatever it may be, is the most important step in the analysis. It will determine what data is required, how they should be collected and how they should be subsequently treated to give the greatest chance of answering the question unambiguously. Collecting data in the vague hope that an interesting question will suggest itself is rarely productive. Remember the aphorism 'Chance favours the prepared mind'.

For the moment we will assume nothing more than an interest in the HPL data in its own right. Does it have any recognizable structure and can this be used to derive a convenient and informative summary of the data? A statistical shorthand for *Table 1* if you like!

The first step is the preparation of a frequency table for the data. The highest and lowest values in *Table 1* are identified and the range they span divided into a convenient number of equally spaced intervals or classes. Ten to 20 intervals are generally recommended, the exact number chosen depending on the type and quantity of data available. Samples of less than 50 might justify the use of less than 10 classes, whilst samples of 1000 or more may well benefit from the use of more than 20 classes.

The intervals should be constructed so as to exclude the possibility of ambiguous allocations. When the data represent counts this difficulty does not arise.

Number of amoebic forms per mm^3	Frequency
1–4	
5–8	
etc.	

Intermediate values between 4 and 5 cannot occur. When dealing with continuous variables an arbitrary dividing point between two adjacent intervals must necessarily form the upper limit of one class and the lower limit of the other.

HPL	Frequency
4·0–5·0	
5·0–6·0	

To which interval do we allocate HPL values of 5·0mg/l? The problem is simple to resolve but does make for some cumbersome terminology. Any measurement involves an element of approximation. An HPL concentration of 5·0mg/l implies a value in the range of 4·95–5·05. Analytical imprecision obliges us to round off our results to one decimal place, e.g. from 4·95 to 5·0. The continuous variable, HPL concentration, is rendered artificially discrete in that no values between 4·9 and 5·0 are recorded. *Class limits* can be simply defined such that values between or equal to those limits are allocated to that class:

Class limits HPL mg/l	Frequency
4·0–4·9	
5·0–5·9	
6·0–6·9	

Although the class limits are unambiguous with reference to the sample data, problems are raised by the discontinuity of the limits when it comes to drawing a histogram of the data (*Fig. 1*).

We are supposed to be dealing with a continuous variable! Although the class limits are all that is generally shown in published frequency tables and histograms, a second set of limits is operational in the construction of histograms. These are the implied class limits or *class boundaries*, which take into account the underlying approximation in the recorded data by expanding the class limits into the next place of decimals, thereby restoring the underlying measurement continuum.

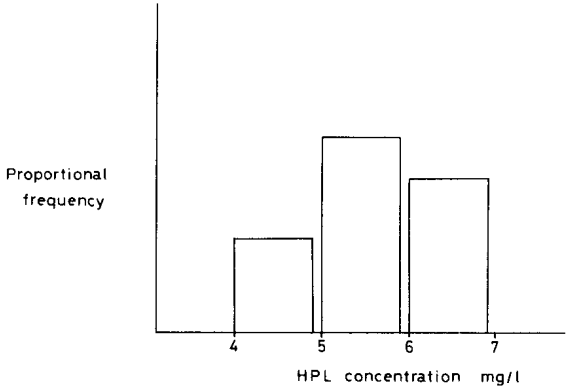

Fig. 1. Histogram defining class limits.

Class limits	Class boundaries	Frequency
4·0–4·9	3·95–4·95	
5·0–5·9	4·95–5·95	
6·0–6·9	5·95–6·95	

Histograms should always be scaled on the class boundaries. If, as is commonly the case, only the class limits are shown, the histogram bars should be drawn a little to the left of the class limits to emphasize the underlying class boundaries (*Fig. 2*).

The frequency table includes two extra columns in *Table 2*. The proportional frequency is obtained by dividing each class frequency by the total number of observations recorded (in the HPL example, 300). This scales the histogram to an overall area of one, considerably simplifying the comparison of histograms based on different numbers of observations. The last column, the cumulative proportional frequency, is obtained by successive addition of the proportional frequencies. The last value will always be 1·00 if the arithmetic is sound. The cumulative proportional frequency has a useful role in assessing the distributional properties of the data, an application referred to in Chapter 4.

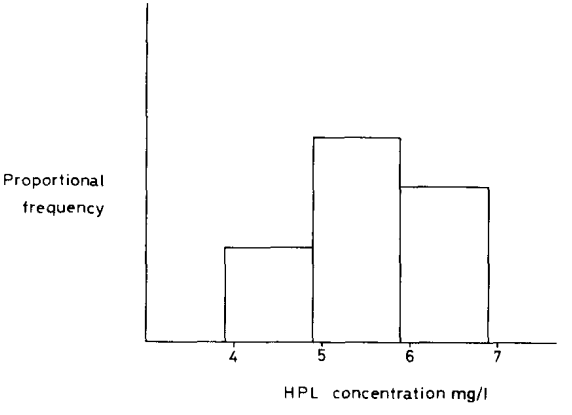

Fig. 2. Histogram defining class boundaries.

If a particular class or classes have very few observations in them they can be pooled. This also serves to 'smooth' out irregularities in the histogram although care should be exercised if valuable information is not to be smoothed into oblivion.

Table 2. Frequency table for HPL values of *Table 1*

Class limits	Class boundaries	Frequency	Proportional frequency	Cumulative proportional frequency
4·0– 4·9	3·95– 4·95	16	0·053	0·053
5·0– 5·9	4·95– 5·95	41	0·137	0·190
6·0– 6·9	5·95– 6·95	61	0·203	0·393
7·0– 7·9	6·95– 7·95	60	0·200	0·593
8·0– 8·9	7·95– 8·95	49	0·163	0·756
9·0– 9·9	8·95– 9·95	30	0·100	0·856
10·0–10·9	9·95–10·95	17	0·057	0·913
11·0–11·9	10·95–11·95	13	0·043	0·956
12·0–12·9	11·95–12·95	8	0·027	0·983
13·0–13·9	12·95–13·95	3	0·010	0·993
14·0–14·9	13·95–14·95	2	0·007	1·000
		$n = 300$	1·000	

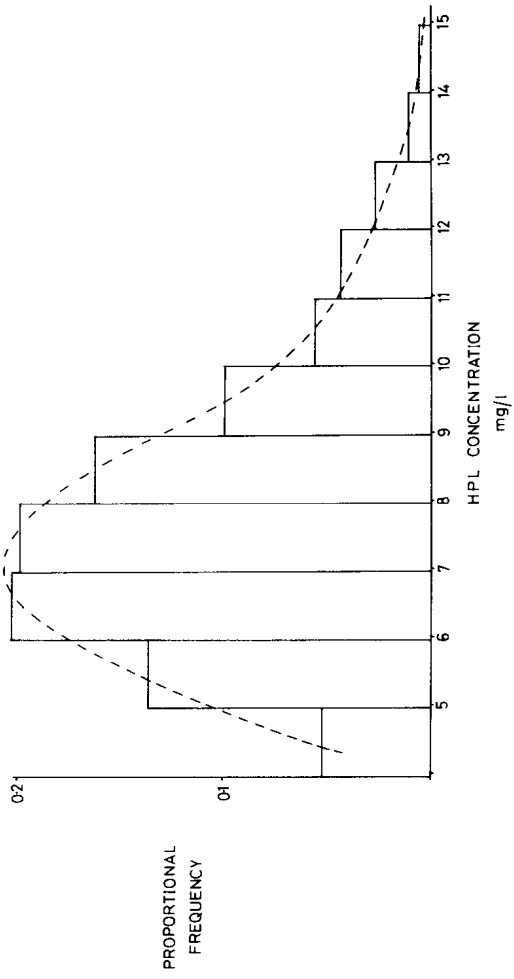

Fig. 3. Histogram of raw sample data from *Table 1*.

A frequency table for the HPL results of *Table 1* is presented in *Table 2* and the corresponding histogram plotted in *Fig. 3*. If the general shape of the histogram is approximated by a curve