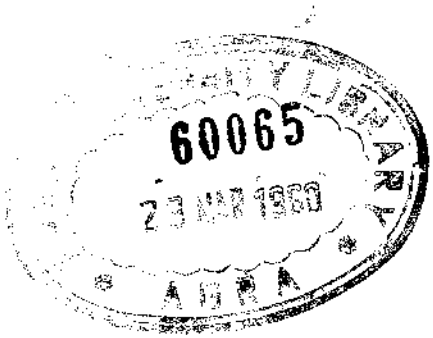
PRINCIPLES OF BIOLOGICAL ASSAY



PRINCIPLES OF
BIOLOGICAL ASSAY

By

C. W. EMMENS
D.Sc., Ph.D.
National Institute for Medical Research, Hampstead, London, N.W.3

With a Foreword by

SIR PERCIVAL HARTLEY
C.B.E., M.C., D.Sc., F.R.S.
formerly Director of Biological Standards, Medical Research Council, The National Institute for Medical Research, Hampstead, N.W.3

LONDON
CHAPMAN & HALL LTD.
37 ESSEX STREET W.C.
1948

First published 1948

THIS BOOK IS PRODUCED IN COMPLETE
CONFORMITY WITH THE AUTHORIZED
ECONOMY STANDARDS

Catalogue No. 354/4

PRINTED AND BOUND IN GREAT BRITAIN BY
WILLIAM CLOWES AND SONS LIMITED, LONDON AND BECCLES

FOREWORD

Biological assay is now a recognised tool for the study of certain properties of living matter. Accuracy in this work is of primary importance whether the object of the assay is the solution of a problem in pure research or the standardisation of a drug. A striving after accuracy has, perhaps, been more obvious in the latter field where biological standardisation has acquired a new importance and direction with the discovery, mainly after the end of the first world war, of a number of new specific remedies which have proved to be of great importance in medicine. Most of these are of biological origin, some are potentially dangerous, many depend on precise dosage for their therapeutic efficiency; they are devoid of characteristic chemical or physical properties which would serve for their identification and assay, and the determination of their potency depends on the reactions they produce in living material. This important group of therapeutic substances includes the antitoxins and tuberculin, the vitamins and insulin, the sex hormones and the active principles of the pituitary gland, the arsphenamines and the heart drugs, heparin and penicillin.

The need for the establishment of reliable methods for the determination of the potency of medicaments of this class is obvious, but progress has frequently been delayed because some of the first attempts were neither well-conceived nor intelligently directed. The ordered progress which is now seen in this field followed the recognition of the inadequacy, and indeed the fallacy, of attempts to define potency in animal units, based on the simple observation of individual reactions produced in animals; the clear demonstration that all methods of biological assay are essentially comparative; and that, for the carrying out of such assays, standard preparations are absolutely necessary. It should never be forgotten that it was due to the sustained and long-continued labours of the Permanent Commission on Biological Standardisation of the Health Organisation of the League of Nations that these standards were provided, firmly established on an international basis, and made freely available throughout the world. In consequence of this achievement in international co-operation, assays carried out in different parts of the world become directly comparable, a common system of unit

notation for the expression of potency is provided, and a sound basis is created for research into the nature, properties, assay, and therapeutic application of these important remedies.

As the author points out, the establishment of these international standards marked a turning point in other developments of biological assay since they permitted statistical methods to be widely applied and fully exploited. The readers of this book will be able to follow the stages of this progress, from the earliest attempts at precise measurement of potency, through the introduction of statistical concepts in devising methods and analysing data; and he will note the great variety of biological reactions which have been exploited for the assay of the different therapeutic and other substances, how this is reflected in the varying designs for assays which have been devised, and in the subsequent treatment of the results. The importance of animal variation in the response to biological stimulus, the bearing of this on the design of assays and on the precision with which they can be applied, the recognition that the errors of biological methods of assay may be large, that the determination of their magnitude is an essential part of the assay itself—all these are part of the story unfolded to the reader of this book.

In the past many of those engaged in this work have too often been content to dispense with the aid of the statistician—or to seek it at the wrong time—in the conduct of their assays and in the interpretation of their data. Indeed, some research workers in biology have a distaste for the symbolism of algebra; but while rigorous logical proof of many important statistical theorems may demand a considerable mathematical training this should be no deterrent to the reader who is less well equipped, for even if he cannot prove rigorously he can verify many important results by using simple arithmetic. One consequence of this useful book should be to encourage all such to increase and perfect their statistical equipment so that in many circumstances they will be able to plan and design their own assays and analyse their own data. With the increasing importance of biological assay in the production of therapeutic substances, of fungicides, insecticides and many other products, in their testing to ensure that official requirements are complied with, and in order to secure and maintain the fullest confidence of those who use them, it is more necessary than ever that those who conduct the assays should be familiar with the many advances in technique which have been made. When to all these considerations is added the fact that research into the chemical nature of these

substances, their purification and their ultimate isolation is inseparably bound up with the determination of the amount present in any material, it is quite clear that biological assay plays a predominant part in whatever field of scientific activity these substances are studied.

The important advances which have followed this close study by biologists and statisticians of problems arising primarily in the field of biological assay are not restricted to that field. Many of the methods here described are of much wider significance, and are applicable in biological fields other than those of standardisation and assay, and even in such disciplines as economics, industrial research, and psychology. One particular direction in which workers in biology are indebted to statisticians arises from their recognition that, very often, the biologist is obliged to work with very small groups; and the help and guidance given in the design of experiments in which the use of small numbers is inevitable, and in the development of statistical methods for the analysis of the data yielded by these methods, have been of very great service and importance.

It may not be out of place, in a book in which statistics play such a predominant part, to emphasise that the least the biologist can do is to bring to his part of the undertaking—the conduct of the assay itself—all the skill, precision and care of which he is capable; but the success of all concerned is bound up, to an extent that statisticians do not always appreciate, with the health and well-being of the experimental animals which are used in practically every assay. The insistence on the provision of healthy stocks of animals is not due to the idea that by this means animal variation can be eliminated, but to the endeavour to ensure that the biological response which is the basis of the assay is not being obscured, and even falsified, by the effects of intercurrent disease. It cannot be over-emphasised that adequate stocks of healthy, well-fed and well-housed animals are as essential to the worker in biological assay as the balance, the microscope and other instruments of precision are to workers in other scientific fields. Moreover, the animal house of a research department and the natural habitat of laboratory animals are different environments, and a knowledge of the natural history of the animals used in assays is not less important than the choice of appropriate and accurate methods of statistical analysis.

It is with the greatest confidence that one can commend this book to the close attention of all interested in this field of scientific activity, and I am grateful to Dr. Emmens for giving me this opportunity of

recording the indebtedness of the Department of Biological Standards at the National Institute for Medical Research to a long succession of colleagues, in other departments, who have come to our aid in the solution of problems which have arisen during the past twenty-five years. Moreover, in the early part of that period we all profited from the close association with professional statisticians who were our colleagues, or shared our domicile, at the Institute at that time; and the close liaison which was then established has happily been maintained. From these fortunate circumstances the literature of biological standardisation and assay has been enriched by many notable contributions and I think we may claim that they have played a small part in winning from a continental writer the description of this country as "the home of modern biometry." Dr. Emmens' book maintains the high standard and tradition of British contributions in this field.

PERCIVAL HARTLEY

The London School of Hygiene and Tropical Medicine,
Keppel Street, W.C. 1.

AUTHOR'S PREFACE

This is not a text-book of statistics, or even of statistical methods of general application to biology. Most of that field has been covered in a number of excellent works in recent years. None of those which have appeared at the time of writing has, however, been particularly concerned with biological assays as such, and very little has been published in book form about the design and analysis of tests made for the purpose of assigning limits within which the potency of preparations may be presumed to fall in comparison with a standard. The present volume is designed to cover the gap.

Although a good deal of elementary statistical procedure is fully described, and some attempt has been made to indicate the reasons behind it, proofs are not given and omissions are many. I have tried to give an account of those statistical methods which are needed in the analysis of biological assays in sufficient detail to make the book self-contained, so as to enable the research worker or routine analyst who is not already an amateur statistician to plan his own tests and analyse them without reference to other manuals. It is to be hoped, nevertheless, that his interest will be sufficiently stimulated to make him wish for a more general acquaintance with a fascinating subject than any one volume can give, particularly a volume written by a non-mathematician. Almost every new method of assaying the potency of biologically-active substances requires statistical treatment which differs at least in small particulars from any of those preceding it. It is therefore impracticable, and would lead to a stultification of the subject, to try to lay down sets of rules for the conduct and analysis of all tests, but it is possible to illustrate by discussion and example some of the basic principles which the design of tests and the treatment of results should follow.

The examples I have used are taken from actual assays, but they have often been modified to illustrate the use of a particular analytical procedure. There are few cases in the literature from which to select examples of the more advanced type, involving restrictions in design and the application of, for instance, covariance analysis. It has sometimes been necessary to impose imaginary conditions on the existing tests, when no suitable example has come to hand, or to use a test designed for a different purpose in the illustration of a

particular method. Thus, the assay of thyrotrophin described in Chapter 6 was not made with a Latin square design, and I have added the dummy restrictions, "cages" and "strains of animal," purely for the purpose of demonstrating the methods of analysis applicable to the Latin square. For this reason I must ask the reader not to take the limits of error found in these examples seriously—they are not always those of the original tests nor those that would be found in practice. Again, in the discussion of assays based on a reaction time, it was convenient to use Mr. D. J. Finney's data from toxicity tests with insects, as an excellent example of the technique of analysis, although these tests do not in fact involve a time factor. It is very difficult and certainly not commendable to invent data to illustrate statistical methods, and therefore far preferable to make the best of existing material.

I am particularly indebted to Mr. A. L. Bacharach, Mr. J. Curry and Dr. C. C. Spicer, who read the manuscript and offered many helpful criticisms and suggestions, and to Professor R. A. Fisher and Dr. F. Yates, also to Messrs. Oliver & Boyd Ltd. of Edinburgh for permission to reprint Tables I, II, III, IV, V and IX from their book *Statistical Tables for Biological, Agricultural and Medical Research*. It is also with much pleasure that I take this opportunity to thank Dr. J. O. Irwin for the many enlightening discussions we have had on statistical matters. Dr. C. I. Bliss has allowed me to reprint various tables from his published works, and I am much in debt to him and his various collaborators for the free use I have made of their material. My thanks are also due to the editors and publishers of *The Biochemical Journal*, *Annals of Applied Biology*, *The Analyst*, *The American Journal of Roentgenology and Radium Therapy* and *The Journal of the American Pharmaceutical Association* for permission to reproduce figures and tables as indicated in the text.

C. W. EMMENS

The National Institute for Medical Research,
Hampstead, London.

CONTENTS

CHAPTER 1

THE APPLICATION OF MATHEMATICS TO BIOLOGICAL MEASUREMENTS

1.1. Introduction

A collection of data contains a certain amount of information on any point. If it has been badly collected it will contain less information than it might have done, and no amount of subsequent juggling with figures will extract more than is inherent in the data.

This concept is axiomatic in modern statistics. It leads to two important conclusions: that the way in which facts are collected for analysis will determine the total amount of information they can yield, and that the complete extraction of the information available depends on using an adequate analytical technique. A badly planned experiment will give results inherently less reliable than a well-planned experiment, whatever is done with them, and an inadequate analysis of the results of the best-planned test will extract only a portion of the information it is capable of conveying.

This is the justification for asking biologists and pharmacologists to concern themselves with the details of statistical techniques. You cannot plan an experiment properly unless you know the methods that will be used in assessing the results, and to use inadequate analytical methods—or, worse still, none at all—in dealing with the results of a test is to prefer guesswork to knowledge. The extent to which guesswork replaces knowledge in the examination of biological assays when a bad analytical method is used is probably not sufficiently appreciated. Few of us would be content to make a chemical extraction by a method which gives 25% of the obtainable yield, yet it is not an exaggeration to say that inefficiency of this order occurs time and time again in the planning and execution of biological assays. That is why we should disagree emphatically with any claim that statistics are never essential in this field. On the contrary, it is quite impossible without using statistics to plan and to analyse the results of an assay in such a way as to obtain the optimum yield of information.

When determining the amount of a substance present in solution

by a chemical assay, it is usually sufficient to repeat the titration once or twice, whereupon by a well-designed method the end point will be found to be substantially constant. Owing to the small errors which are usually encountered in such determinations in inorganic and even in organic chemistry, chemical and biological workers found it unnecessary to estimate the error of their titrations or analyses, or, at most, to form more than an approximate idea of their magnitude.

In some of the even finer scientific measurements made in physics, particularly in optics, where the precision of an answer is usually greater than that of chemical analysis, it was, curiously enough, found necessary to investigate methods for determining the error with which very accurate determinations were made. Such computations of error led to the development of a theory which described the "normal curve of error" encountered in making repeated determinations of such refined observations as the length of the standard metre or the weight of the standard kilogram, when expressed in the units of some other system of measures. These early statistical methods were, of course, available to the chemist if he found them necessary and was aware of their existence.

It was not until comparatively recently that any serious attempts were made to assay chemical or pharmacological substances by other than chemical means. Now that widening researches into a variety of fields, such as the vitamins and hormones, have rendered necessary the development of assay techniques using biological material, there has occurred within the past two decades a rapidly growing study of statistical methods for attacking the problems that arise. The range of variation encountered in biological assay is very large compared with that to which the earlier statistical techniques were applied, and the conditions under which assays must often be conducted make possible interference from various sources of a type which was hardly ever encountered, or at any rate recognised, in chemical or physical measurements. Thus, statistical techniques which were designed for refining procedures which themselves had a high inherent precision have had to be enlarged and developed for dealing with types of measurement not only of a low inherent precision but clearly subject to various types of error not previously investigated.

The earlier workers in the field of biological assay were, for the most part, unacquainted with those statistical procedures which were at that time available to them, and often peculiarly blind to the

degree of animal variation which was staring them in the face. It is not long since biologists competent in other fields were capable of comparing the effect of a dose of one drug on one animal with that of another drug on a second animal, and of drawing conclusions as to the relative potencies of the two from such meagre data. As it was gradually realised that comparisons of this nature were not only subject to relatively enormous errors but also that the magnitude of the errors was unknown, attempts were made to better the situation by using groups of animals instead of single animals, and by constructing curves relating the various doses of a drug to the average response of groups of animals, each receiving the same dose. This procedure reduced the error of determinations of relative potency, but left its formulation in an unsatisfactory state.

The next step was for biologists to make the attempt both to understand the elements of the necessary mathematics and, at the same time, to call in the aid of professional statisticians to examine the type of problem with which they were dealing and to evolve methods for expressing relative potencies and their errors in as simple a manner as possible. Unfortunately, professional statistical help was even then dispensed with to a deplorable extent, with the result that at the present time many official and unofficial procedures for the determination of relative potencies are based on invalid assumptions. The situation is fortunately undergoing a rapid change for the better. This change involves as an integral part of the reformation of methods of biological assay the recognition that it is not enough to ask the statistician to help in elucidating the results of biological work after they have been completed. His assistance in planning the original research (or an adequate understanding of the problem on the part of the experimenter), so that it shall be most easily handled by the statistical methods now available, is essential for eliminating wasteful or useless work in the laboratory and for attaining a maximum of precision.

1.2. Animal Units

The earliest attempts at defining relative potencies of drugs—by which is meant here any biologically active substance or poison and even other agents such as heat or radiation—were in terms of animal units. Thus, the unit of androgenic hormone, the substance responsible for the development of the typical secondary sexual characters of the male, was that amount which, when injected intramuscularly into capons, would produce a certain minimal

degree of comb growth in a group of the birds. This was the "capon unit." There were many other such units and their unsatisfactoriness was soon apparent. They were unsatisfactory for a variety of reasons. In the first place it was very difficult so to adjust dosages that the response obtained to a given dose was sufficiently near to the unit as defined to satisfy the necessary criteria, and various attempts at the establishment of some kind of dose-response curve in terms of animal units were found to be unreliable. Secondly, the variation which biological material usually exhibits, not only between animals or plants or preparations from them, but from time to time in the same organism or stock, rendered the unit most unstable.

Tests at present specified in the U.S. Pharmacopeia for the determination of potencies of some of the vitamins exemplify the best that can be done by the use of such animal units. A standard substance and the unknown are compared at the same time and the procedure stipulates that the response to the unknown must be at least equal to that of the standard. It is therefore necessary for the substance tested to attain a certain minimal potency in order to pass the test. This introduces the concept of using a standard by which to assess the potencies of other substances in the test, and the setting up of standards has marked a turning-point in the progress of biological assay and has enabled the more powerful statistical methods now available to be more fully exploited.

1.3. International Standards

A large number of international standard preparations has now been set up. They are intended to be stable yardsticks against which the potency of other substances may be measured at any time. They are kept under conditions designed to preserve their activity unchanged and distributed from such centres as the National Institute for Medical Research in London, while careful checks on their stability are made at frequent intervals. Standards now exist in crystalline form for many of the sex hormones, for soluble insulin, and some of the vitamins, and in various impure forms for other vitamins, hormones, anti-toxins and other pharmacologically active substances. It is now a common practice, in order to conserve such international standards, for laboratories to set up their own sub-standards, which have been carefully assayed in comparison with the international standard preparation. In the investigation of substances for which no international standard

preparation is available, laboratories also make a practice of setting up their own provisional standards for use in research.

When a standard preparation is available and has been demonstrated as far as is possible to be stable, any new substance may be compared with it in order to determine whether it has the same action as the standard, and, if so, to express its potency in terms of the standard. One of the first ways in which progress was made in the use of standard preparations was by the establishment of a dose-response curve determined, for instance, by injecting several different doses of the standard preparation each to a group of animals and relating response to dose graphically. The potency of another substance having the same pharmacological action was then related to that of the standard by injecting one or more groups of animals with known doses of it, and reading backward from the dose-response curve the doses of the standard that would be supposed to produce the same response as each dose of the preparation under test. Variation in response which may occur from time to time was often ignored in these procedures, or an attempt was made to allow for it by giving further groups of animals one or more doses of the standard at the same time as the unknown was administered and assessing roughly whether or not the position of the dose-response curve remained constant.

1.4. Curve-fitting

All kinds of curves were fitted to the results of assays. More often than not, curves were fitted by eye and no equations were determined for them. When equations were determined, they were most frequently simple parabolas relating dose directly to response and were of limited use in the estimation of error. It was apparent that the calculation of error could not approach exactness unless the dose-response data were fitted by relatively few types of standard curve, which could be investigated statistically and for which suitable mathematical formulae to be used in the estimation of relative potencies and of their errors could be computed. The fact that the curves which came into general use were not necessarily those by which the data could be fitted with the greatest exactitude is of minor importance compared with the mathematical suitability of the types of curve finally adopted. Any series of results to be used in the computation of a standard dose-response curve may in practice be fitted by an infinity of mathematical functions, many of which, although demonstrably inappropriate at the extreme ranges

of response, are as good a fit as can be expected over that range which will be used in practice. Faced, therefore, with such a choice of functions, it is logical to adopt those which are most easily handled in computing relative potencies and in estimating the precision with which these potencies are determined, as long as they do not grossly violate any theoretical considerations as to the probable nature of the dose-response relationship.

1.5. Increasing the accuracy of tests

Having settled these difficulties, that a standard preparation must be used and that the response to it shall be related to the dose by the most convenient method of graduation which fits the facts, the relatively rapid strides which have been made in increasing the accuracy of assays became possible. In the following chapters we shall see how methods have been developed for eliminating the effect of many sources of variation inherent in biological assays and for so planning the comparison of a substance with the standard preparation that the assay shall be self-contained. This means that measures shall be available of the validity of comparing the two substances, as well as estimates of the limits within which the potency of the unknown may be judged to fall in comparison with the standard, to any desired degree of probability. From the internal evidence of a well-planned assay it is also possible to make adjustments for various other concomitant factors by which the response may be partially conditioned and to eliminate their effects in the estimation of potency.

Side by side with these refinements of statistical technique have been developed refinements in the conduct of the assays themselves. The increase in precision which may result from the adoption of modifications in the conduct of assays is, of course, measured by statistical methods. They serve to eliminate unnecessary procedures which on *a priori* grounds might be thought to add to the accuracy of the tests, but may turn out upon critical examination to do no such thing. It is typical of biological assay that homogeneity of the test material is sought, although much of its heterogeneity may be eliminated by adequate statistical control. Thus, the use of pure lines of animals has served to eliminate part of the variation found within less homogeneous animal stocks, and the running of the test under rigidly defined conditions will usually reduce the variability of the results to a worth-while extent. This narrowing of the conditions under which assays are conducted, adopted for

the purpose of reducing error as far as is conveniently possible, follows on our knowledge that the substances which are being compared in such tests are of a well-defined type, known to possess certain pharmacological or other properties. In these tests the animal or plant preparation is, in fact, used as a test-tube and would not normally be so used unless a good deal were already known about the properties of the substances being compared. In other words, the test is run upon a narrow inductive basis.

There are many statistical methods, some of them identical with those used in biological assay, which are appropriate to the testing of materials and the comparison of their effects when a broader inductive basis is desired. When investigating the properties of substances in a relatively new field of work, it may be undesirable to confine one's experiments to a very homogeneous population of animals, since misleading results may be obtained which, however true for the particular stocks and test conditions utilised, may not be of wide application to other even closely related biological material. Although in assaying the potencies of particular vitamin preparations, it may be desirable to use only the males of a particular breed of rats, it would be a mistake to base either qualitative or quantitative conclusions about the action of the vitamin in the animal kingdom in general upon the results obtained from such homogeneous material, for it may be discovered that the relationships utilised in assaying with one given type of animal do not hold even for other stocks of the same species. It is useful, therefore, to distinguish at the outset between the frequent desirability of rigid control of conditions in a highly developed assay technique from the wider type of experimentation that will usually be preferable in the investigation of the properties of new and little-known substances.

CHAPTER 2

MEANS, VARIANCES AND DEGREES OF FREEDOM

2.1. The normal distribution

It we make repeated attempts to measure a quantity as accurately as possible we usually find that the results are scattered about their average or mean value in a typical manner. The measurements tend to cluster most closely around the mean and to thin out as we depart from it. The distribution which such measurements usually exhibit is called the *normal distribution* or *normal curve of error*. The use of the word "error" was introduced in describing the curve because the earlier applications of it were concerned with the errors of measurement of such physical quantities as the height of a building or the length of standard measures. The curve is bell-shaped and has the general formula:

$$p = \frac{1}{\sigma\sqrt{2\pi}} e^{-Y^2/2\sigma^2}$$

where Y is the variable being measured and p the frequency at a given value of Y. Fig. 2.1 illustrates the normal distribution, and it will be seen that the curve reaches a maximum at the mean, falling off rapidly on either side of it, and has two points of inflection, one on each side of the mean. Table 2.1 gives the ordinates of the curve. These points of inflection (at $Y = \pm\sigma$) are the points where the curve is steepest and below them the curve falls off less and less steeply, so that its two arms slowly approach the base-line, but never reach it. Approximately two-thirds of the total area enclosed by the curve is contained within two vertical lines drawn from the two points of inflection, and these lines mark off along the base-line the natural units in which the curve is measured and are thus by definition situated at plus or minus one *standard deviation* ($\pm\sigma$) on each side of the mean.

These conditions would only be fulfilled in practice if a very large number of observations were made from a normally distributed population, when the observed curve takes very nearly its theoretical mathematical form. (A population is a theoretical

infinity of possible observations from which we actually take a sample.) When relatively few observations are made, their distribution, even when they are drawn from a population which itself is normally distributed, copies the normal curve of error in only a sketchy form.

It has been found that this type of curve describes not only the distribution of errors encountered in attempts at accurate physical measurements of various kinds, but that in a large number of cases it equally well describes the distribution of measurements of various biochemical and biological phenomena. Thus, measurements of

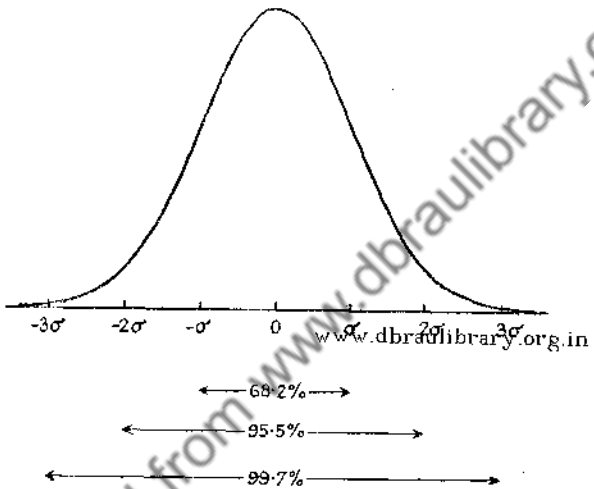


FIG. 2.1. The normal curve of error, calibrated in standard deviations, measured from the mean. The arrows indicate the percentage of all observations which fall within the limits $\pm\sigma$, $\pm 2\sigma$ and $\pm 3\sigma$.

such quantities as the heights or weights of populations of animals, the length of stamens of a given type of plant and the amount of haemoglobin in blood samples are frequently found to be normally distributed. This is true also of the measurements of the response of animal or plant preparations in the groups which are used for biological assays. It will be appreciated that in a large number of biological tests there will never have been sufficiently extensive investigation for a really critical analysis of just how well the distribution of responses is fitted by the normal curve of error, but it is a fortunate fact that the distribution of *means* of samples from even quite abnormally distributed populations tends rapidly to normality as the sample size increases.

TABLE 2.1

Ordinates of the Normal Distribution

x	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09	1	2	3	4	5
0·0	·3989	·3989	·3989	·3988	·3986	·3984	·3982	·3980	·3977	·3973	0	0	–1	–1	–1
0·1	·3970	·3965	·3961	·3956	·3951	·3945	·3939	·3932	·3925	·3918	–1	–1	–2	–2	–3
0·2	·3910	·3902	·3894	·3885	·3876	·3867	·3857	·3847	·3836	·3825	–1	–2	–3	–4	–5
0·3	·3814	·3802	·3790	·3778	·3765	·3752	·3739	·3725	·3712	·3697	–1	–3	–4	–5	–6
0·4	·3683	·3668	·3653	·3637	·3621	·3605	·3589	·3572	·3555	·3538	–2	–3	–5	–6	–8
0·5	·3521	·3503	·3485	·3467	·3448	·3429	·3410	·3391	·3372	·3352	–2	–4	–6	–8	–9
0·6	·3332	·3312	·3292	·3271	·3251	·3230	·3209	·3187	·3166	·3144	–2	–4	–6	–8	–10
0·7	·3123	·3101	·3079	·3056	·3034	·3011	·2989	·2966	·2943	·2920	–2	–5	–7	–9	–11
0·8	·2897	·2874	·2850	·2827	·2803	·2780	·2756	·2732	·2709	·2685	–2	–5	–7	–9	–12
0·9	·2661	·2637	·2613	·2589	·2565	·2541	·2516	·2492	·2468	·2444	–2	–5	–7	–10	–12
1·0	·2420	·2396	·2371	·2347	·2323	·2299	·2275	·2251	·2227	·2203	–2	–5	–7	–10	–12
1·1	·2179	·2155	·2131	·2107	·2803	·2059	·2036	·2012	·1989	·1965	–2	–5	–7	–10	–12
1·2	·1942	·1919	·1895	·1872	·1849	·1826	·1804	·1781	·1758	·1736	–2	–5	–7	–9	–11
1·3	·1714	·1691	·1669	·1647	·1626	·1604	·1582	·1561	·1539	·1518	–2	–4	–7	–9	–11
1·4	·1497	·1476	·1456	·1435	·1415	·1394	·1374	·1354	·1334	·1315	–2	–4	–6	–8	–10
1·5	·1295	·1276	·1257	·1238	·1219	·1200	·1182	·1163	·1145	·1127	–2	–4	–6	–7	–9
1·6	·1109	·1092	·1074	·1057	·1040	·1023	·1006	·0989	·0973	·0957	–2	–3	–5	–7	–8
1·7	·0940	·0925	·0909	·0893	·0878	·0863	·0848	·0833	·0818	·0804	–2	–3	–5	–6	–8
1·8	·0790	·0775	·0761	·0748	·0734	·0721	·0707	·0694	·0681	·0669	–1	–3	–4	–5	–7
1·9	·0656	·0644	·0632	·0620	·0608	·0596	·0584	·0573	·0562	·0551	–1	–2	–4	–5	–6

2·0	·0540	·0529	·0519	·0508	·0498	·0488	·0478	·0468	·0459	·0449	−1	−2	−3	−4	−5
2·1	·0440	·0431	·0422	·0413	·0404	·0396	·0387	·0379	·0371	·0363	−1	−2	−3	−3	−4
2·2	·0355	·0347	·0339	·0332	·0325	·0317	·0310	·0303	·0297	·0290	−1	−1	−2	−3	−4
2·3	·0283	·0277	·0270	·0264	·0258	·0252	·0246	·0241	·0235	·0229	−1	−1	−2	−2	−3
2·4	·0224	·0219	·0213	·0208	·0203	·0198	·0194	·0189	·0184	·0180	0	−1	−1	−2	−2
2·5	·0175	·0171	·0167	·0163	·0158	·0154	·0151	·0147	·0143	·0139	0	−1	−1	−2	−2
2·6	·0136	·0132	·0129	·0126	·0122	·0119	·0116	·0113	·0110	·0107	0	−1	−1	−1	−2
2·7	·0104	·0101	·0099	·0096	·0093	·0091	·0088	·0086	·0084	·0081	0	−1	−1	−1	−1
2·7	·0079	·0077	·0075	·0073	·0071	·0069	·0067	·0065	·0063	·0061	0	0	−1	−1	−1
2·9	·0060	·0058	·0056	·0055	·0053	·0051	·0050	·0048	·0047	·0046	0	0	0	−1	−1

	·0	·1	·2	·3	·4	·5	·6	·7	·8	·9
3·0	·0044	·0033	·0024	·0017	·0012	·0009	·0006	·0004	·0003	·000

Table 2.1 is reprinted from Table II of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

2.2. Elementary statistical definitions

If Y is any individual observation, the sum of all observations is denoted by SY. The mean of these observations is $\frac{SY}{n}$, where n is the total number of observations which have been made. This mean is denoted by $\bar{Y}$, and we thus have the relationship:

$$\bar{Y}=\frac{SY}{n}$$

The symbol S always implies the sum of all the quantities represented by the symbol, or group of symbols, immediately following it, or of all the symbols contained in the brackets following S. Thus $S(Y-\bar{Y})$ would imply the sum of all the differences of Y and $\bar{Y}$. The individual deviations of each separate Y from the mean, which will be $Y_1-\bar{Y}$, $Y_2-\bar{Y}$ . . . $Y_n-\bar{Y}$, are written as y_1, y_2 . . . y_n, so that in general $y=Y-\bar{Y}$. The sum of all ys is zero, i.e. $Sy=0$.

One of the first needs that arise in computation is that of representing the splay which various observations exhibit about their mean. The direct addition of all their individual deviations results, as we have seen, in zero sum. It would be possible by giving all the deviations a positive sign to measure the average deviation, whether positive or negative, and this statistic has before now been utilised. Such a way of representing the splay does not lead to very useful statistical methods. There are, on the other hand, strong theoretical reasons for utilising the powers of y for describing the properties of the normal and other distributions and for examining samples supposed to have been drawn from a normally distributed population. Thus, while the mean is so determined that $Sy=0$, the splay around the mean is measured by Sy^2, which is always a positive quantity. Other powers of y, i.e. y^3 or y^4, are used in measuring further properties of distribution.

The use of y^2 in measuring the scatter of observations leads naturally to the employment of the standard deviation, σ. The relation between the sum of squares of the deviations and the standard deviation, usually denoted by the lower case Greek letter σ employed above, is such that:

$$\sigma^2=\frac{Sy^2}{n-1}$$

The estimate of σ that we make in practice from our observations is denoted by s. The standard deviation is thus a kind of average

of y^2s, but the divisor for determining this average is $n-1$ and not n. It is difficult without going into further mathematics to explain the reasons for dividing by $n-1$ instead of n when determining standard deviations, and in cases where n is large the difference between dividing by n and $n-1$ is clearly insignificant. This is not so when n is small, and it may be shown that the difference just compensates on the average for the inherent inaccuracy of small samples, and that using n as a divisor leads on the average to an underestimation of σ.

The square of the standard deviation, σ^2, which appears in the formula above, is called the *variance* or *mean square*, usually denoted also by the letter V. Thus, if s from one group of observations is twice that from another, the corresponding variances will be in the ratio 4 : 1. The reciprocal of the variance gives by definition an estimate of the *amount of information* supplied by any one observation in a group and is used as a measure of the relative precision of observations in different groups. Thus, on the average, an observation from a group with a standard deviation which is twice that of a second group will give a quarter of the information supplied by an observation from the second group.

2.3. The variance of a mean

In most elementary text-books of statistics there will be found numerical examples demonstrating that the variance of the mean of a set of observations is $\frac{1}{n}$ of that of any individual observation, where n is the number of observations in the group. This is equivalent to saying that there is available n times the amount of information about a mean as there is about any individual observation in the set. Thus

$$V\bar{Y} = s_{\bar{Y}}^2 = \frac{Sy^2}{n(n-1)}$$

It follows, therefore, that the *standard error* of a mean of n observations is $\frac{1}{\sqrt{n}}$ times the standard deviation of the group. If, for instance, 100 observations are used in determining the mean, the standard error of this mean will be 1/10th of the standard deviation for the whole population, and on the average a series of means of 100 such observations will have 1/10th of the range of the corresponding individual observations. The term *standard error* distinguishes

TABLE 2.2

The Normal Distribution

P	·00	·01	·02	·03	·04	·05	·06	·07	·08	·09
·0	∞	2·575829	2·326348	2·170090	2·053749	1·959964	1·880794	1·811911	1·750686	1·695398
·1	1·644854	1·598193	1·554774	1·514102	1·475791	1·439521	1·405072	1·372204	1·340755	1·310579
·2	1·281552	1·253565	1·226528	1·200359	1·174987	1·150349	1·126391	1·103063	1·080319	1·058122
·3	1·036433	1·015222	·994458	·974114	·954165	·934589	·915365	·896473	·877896	·859617
·4	·841621	·823894	·806421	·789192	·772193	·755415	·738847	·722479	·706303	·690309
·5	·674490	·658838	·643345	·628006	·612813	·597760	·582841	·568051	·553385	·538836
·6	·524401	·510073	·495850	·481727	·467699	·453762	·439913	·426148	·412463	·398855
·7	·385320	·371856	·358459	·345125	·331853	·318639	·305481	·292375	·279319	·266311
·8	·253347	·240426	·227545	·214702	·201893	·189118	·176374	·163658	·150969	·138304
·9	·125661	·113039	·100434	·087845	·075270	·062707	·050154	·037608	·025069	·012533

P	·001	·000,1	·000,01	·000,001	·000,000,1	·000,000,01	·000,000,001
x	3·29053	3·89059	4·41717	4·89164	5·32672	5·73073	6·10941

The value of P for each entry is found by adding the column heading to the value in the left-hand margin. The corresponding value of x is the deviation such that the probability of an observation falling outside the range from $-x$ to $+x$ is P. For example, $P=03$ for $x=2{\cdot}170090$; so that 3 per cent. of normally distributed values will have positive or negative deviations exceeding the standard deviation in the ratio 2·170090 at least.

The table of probits (Table 14.3) provides a more extensive (P, x) table. The probit table refers to a single tail of the distribution, and the P's derived from that table must therefore be multiplied by 2 to bring them into line with the P's of Table 2.2.

Table 2.2 is reprinted from Table I of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

the square root of the variance of a mean from the *standard deviation* of individual observations.

2.4. Small samples

When describing the normal distribution, it was seen that approximately two-thirds of all the observations will fall within the range $\pm\sigma$ about the mean. It may be shown that approximately 21/22 of all observations will fall within the range $\pm 2\sigma$, i.e. a total range of 4σ with the mean as its centre. Table 2.2 shows the proportion of observations which fall within various ranges by steps of 0·01 (i.e. 1%). It will be seen from the table that less than one observation in 10,000 will fall outside the range of $\pm 4\sigma$ in a normally distributed population. The above describes the conditions found to hold when large numbers of observations, of the order of hundreds or thousands, are examined. When a small sample is dealt with, the mean and an estimate of its variance are calculated by the usual methods, but unless the sample exhibits extreme departure from the normal distribution, it is impossible to judge whether it is drawn from a population in itself normally distributed. In the absence of evidence to the contrary, it is assumed that the normal distribution holds and the statistical methods applicable to small samples from the normal distribution are used. It has fortunately been demonstrated that in many of the tests which will be developed in later chapters, minor departures from normality of distribution —and in some of them quite large departures—do not seriously affect the validity of the statistical conclusions drawn.

2.5. Calculation of means and standard deviations

In Table 2.3, Column 1, are listed 20 measurements of the uterine weight of members of a group of young female rats. The weights are taken to the nearest mgm. These weights, which vary from 9 to 32 mgm., are an example of the large splay so frequently encountered in biological measurements, and it is clear that the chance that any one of them is sufficiently representative of the whole group for the purposes of comparison with other groups is not large.

The mean is determined by summing the 20 weights and dividing by 20, the number in the group. It is found to be exactly 21. The deviations from the mean are listed in the second column in the Table. The deviations for numbers falling below 21 are, of course, negative. The sum of these deviations is, by definition,

TABLE 2.3

CALCULATION OF A MEAN AND STANDARD DEVIATION

Y	y	y^2
9	−12	144
14	−7	49
15	−6	36
15	−6	36
16	−5	25
18	−3	9
18	−3	9
19	−2	4
19	−2	4
20	−1	1
21	0	0
22	1	1
22	1	1
24	3	9
24	3	9
26	5	25
27	6	36
29	8	64
30	9	81
32	11	121
Totals 420	0	664
Means 21	0	—

$$s^2=\frac{664}{19}=34{\cdot}9473 \qquad s=5{\cdot}91$$

$$s_{\bar{Y}}^2=\frac{s^2}{20}=1{\cdot}74737 \qquad s_{\bar{Y}}=1{\cdot}32$$

zero. To estimate the standard deviation we sum the squares of the individual deviations. These squares are listed in the third column of the Table, and since all squares are positive in sign, the sum of these squares is a positive quantity, namely 664. This quantity is called, simply, the *sum of squares*. In order to derive the value of the standard deviation from this sum we divide by $n-1$, which gives us the variance, 34·9473.

It is in practice unnecessary to compute the actual separate deviations given in column 2, or their corresponding squares, since the sum of squares can be calculated very rapidly and with less chance of a slip in the calculations from the following relationship:

$$Sy^2=SY^2-n\bar{Y}^2=SY^2-\bar{Y}T=SY^2-\frac{T^2}{n}$$

where T is the total of all observations. We use whichever form of the relationship is most convenient, and where, say, $\bar{Y}T$ is written in the following pages, it is often easier to divide T^2 by n. When using this relationship we do not need to know the individual values of the deviations, and although we have to deal with the squares of larger numbers, we avoid the task of determining each individual deviation from the mean. Since the mean will rarely be a whole number, as in this sample, this second method of calculation is usually the easier. In the present example, $SY^2=9{,}484$ and $\bar{Y}T=8{,}820$, their difference being 664, as it should. The variance computed above is the variance of the individual observations and the standard deviation applicable to these observations is thus $\sqrt{34{\cdot}9473}$, or 5·91.

We can roughly check the absence of a slip in the calculations by remembering that the variance is computed on the assumption that the various estimates with which we are dealing are normally distributed and that approximately two-thirds of them will fall within the range of $\pm s$ of the mean and that approximately 21/22 will fall within the range of $\pm 2s$. Glancing back at the second column of Table 2.3 we see that in fact 12 observations fall within the range $\pm 5{\cdot}91$ and 19 fall within the range $\pm 11{\cdot}82$, and there is thus no reason to suspect any error in our calculations. Note that this procedure is not a check on the accuracy of our assumption that the distribution is normal, but is a quick check on the value found for s. Duplicate calculation is essential for a full check.

The standard error of the mean uterine weight in the group of 20 is given by the formula:

$$V\bar{Y}=s_{\bar{Y}}^2=\frac{s^2}{n}=\frac{Sy^2}{n(n-1)}$$

a calculation based, it will be recalled, on the fact that we have n times the information about the mean that we have about any of the individual observations contributing to it. The standard error of the mean in this instance is thus:

$$\frac{\sqrt{664}}{20\times 19}=1{\cdot}32$$

What does the standard error tell us about the mean? It is, of course, an index of its variability and a measure from which we can compute the chances that the mean we have observed differs from any other assigned value.

TABLE 2.4

DISTRIBUTION OF t

	Probability												
n	·9	·8	·7	·6	·5	·4	·3	·2	·1	·05	·02	·01	·001
1	·158	·325	·510	·727	1·000	1·376	1·963	3·078	6·314	12·706	31·821	63·657	636·619
2	·142	·289	·445	·617	·816	1·061	1·386	1·886	2·920	4·303	6·965	9·925	31·598
3	·137	·277	·424	·584	·765	·978	1·250	1·638	2·353	3·182	4·541	5·841	12·941
4	·134	·271	·414	·569	·741	·941	1·190	1·533	2·132	2·776	3·747	4·604	8·610
5	·132	·267	·408	·559	·727	·920	1·156	1·476	2·015	2·571	3·365	4·032	6·859
6	·131	·265	·404	·553	·718	·906	1·134	1·440	1·943	2·447	3·143	3·707	5·959
7	·130	·263	·402	·549	·711	·896	1·119	1·415	1·895	2·365	2·998	3·499	5·405
8	·130	·262	·399	·546	·706	·889	1·108	1·397	1·860	2·306	2·896	3·355	5·041
9	·129	·261	·398	·543	·703	·883	1·100	1·383	1·833	2·262	2·821	3·250	4·781
10	·129	·260	·397	·542	·700	·879	1·093	1·372	1·812	2·228	2·764	3·169	4·587
11	·129	·260	·396	·540	·697	·876	1·088	1·363	1·796	2·201	2·718	3·106	4·437
12	·128	·259	·395	·539	·695	·873	1·083	1·356	1·782	2·179	2·681	3·055	4·318
13	·128	·259	·394	·538	·694	·870	1·079	1·350	1·771	2·160	2·650	3·012	4·221
14	·128	·258	·393	·537	·692	·868	1·076	1·345	1·761	2·145	2·624	2·977	4·140
15	·128	·258	·393	·536	·691	·866	1·074	1·341	1·753	2·131	2·602	2·947	4·073
16	·128	·258	·392	·535	·690	·865	1·071	1·337	1·746	2·120	2·583	2·921	4·015
17	·128	·257	·392	·534	·689	·863	1·069	1·333	1·740	2·110	2·567	2·898	3·965
18	·127	·257	·392	·534	·688	·862	1·067	1·330	1·734	2·101	2·552	2·872	3·922
19	·127	·257	·391	·533	·688	·861	1·066	1·328	1·729	2·093	2·539	2·861	3·883
20	·127	·257	·391	·533	·687	·860	1·064	1·325	1·725	2·086	2·528	2·845	3·850

21	·127	·257	·391	·532	·686	·859	1·063	1·323	1·721	2·080	2·518	2·831	3·819
22	·127	·256	·390	·532	·686	·858	1·061	1·321	1·717	2·074	2·508	2·819	3·792
23	·127	·256	·390	·532	·685	·858	1·060	1·319	1·714	2·069	2·500	2·807	3·767
24	·127	·256	·390	·531	·685	·857	1·059	1·318	1·711	2·064	2·492	2·797	3·745
25	·127	·256	·390	·531	·684	·956	1·058	1·316	1·708	2·060	2·485	2·787	3·725
26	·127	·256	·390	·531	·684	·856	1·058	1·315	1·706	2·056	2·479	2·779	3·707
27	·127	·256	·389	·531	·684	·855	1·057	1·314	1·703	2·052	2·473	2·771	3·690
28	·127	·256	·389	·530	·683	·855	1·056	1·313	1·701	2·048	2·467	2·763	3·674
29	·127	·256	·389	·530	·683	·854	1·055	1·311	1·699	2·045	2·462	2·756	3·659
30	·127	·256	·389	·530	·683	·854	1·055	1·310	1·697	2·042	2·457	2·750	3·646
40	·126	·255	·388	·529	·681	·851	1·050	1·303	1·684	2·021	2·423	2·704	3·551
60	·126	·254	·387	·527	·679	·848	1·046	1·296	1·671	2·000	2·390	2·660	3·460
120	·126	·254	·386	·526	·677	·845	1·041	1·289	1·658	1·980	2·358	2·617	3·373
8	·126	·253	·385	·524	·674	·842	1·036	1·282	1·645	1·960	2·326	2·576	3·291

Table 2.4 is reprinted from Table III of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

2.6. The *t* test

A function, *t*, first investigated by "Student" (W. S. Gosset), has been tabulated for this purpose and is such that

$$t=\frac{\bar{Y}-m}{s\bar{Y}},$$

where m is a value known or postulated for the mean of the population. The values of t when the number of observations approaches infinity are those in Table 2.2. Table 2.4 gives values of t found from samples of a normally distributed population at various levels of probability, for numbers of observations from two upwards. Thus when the number of observations is 20, as at present, and the number of degrees of freedom (see below) is 19, there is a 5% chance that t equals or exceeds 2·093 and a 1% chance that it exceeds 2·861. We may therefore assert that the chance is 1 in 20 that the sample has been drawn from a population the mean of which lies outside the limits $21\pm(1{\cdot}32\times2{\cdot}093)$, or 18·24 to 23·76, and is 1 in 100 that it lies outside the limits $21\pm(1{\cdot}32\times2{\cdot}861)$, or 17·22 to 24·78. Note that our best estimate of the value of the mean is the actual value we find, but that it is an estimate to which we can assign certain specific limits of accuracy by the use of the standard error. These 5% or 1% chances are alternatively expressed by saying that P, the probability of equalling or exceeding a given value of t, is 0·05 or 0·01.

2.7. The coefficient of variation

A statistic which is quite frequently used, and which is calculated by dividing the standard deviation by the mean and multiplying by 100, is the *coefficient of variation*. This coefficient is the standard deviation expressed as a percentage of the mean. Its value depends on the magnitude of both statistics, and thus the same variance attached to a big mean will yield a smaller coefficient of variation than if it is attached to a smaller mean. The coefficient of variation of the series above is $\frac{5{\cdot}91}{21}\times100$, or 28·1%. If we were to add 1,000 to every figure in the first column of Table 2.3, the mean would be 1,021, the standard deviation would remain the same, but the coefficient of variation would fall to $\frac{5{\cdot}91\times100}{1{,}021}$, or 0·58% approx. Such facts are important if, in the determination of means and standard errors, we use coded figures.

2.8. Coding

We code figures for rapid computation when the crude data are unwieldy. Thus, in a series which ranged from 1,009 to 1,032, we would code before computation by subtracting 1,000 from each number—a procedure which can be followed with no risk of error—and treat the data just as in the foregoing. We must remember, however, to add 1,000 on again after we have completed our calculations and also remember that coding of this type does not affect the value of the variance. In other instances we may code by rounding to the nearest whole number if our original data appear to be unnecessarily detailed, and in this instance coding will have a negligible effect on both the mean and the variance. If in the process of coding we have performed any more complicated procedure, such as dividing, or dividing and then subtracting a constant number from each observation, we must at the end of our calculations repeat the coding procedure in strictly reverse order on the statistics we have derived, in order to finish up with means and variances of the correct magnitude. Thus, if we divided each number by 10 and then subtracted 80, the true mean would be arrived at by adding 80 and then multiplying by 10, while the variance would have to be corrected by multiplying by 100, which is the same as multiplying the standard deviation by 10.

2.9. Degrees of freedom

A set of observations such as that in Table 2.3 is called a *group*, *class* or *array*. This array contains 20 items and there are 19 *independent comparisons* or *degrees of freedom* which exist within the array. The first 10 of these comparisons may, if we so desire, be made between successive pairs of items, $Y_1 - Y_2$; $Y_3 - Y_4$ . . . $Y_{19} - Y_{20}$. The remaining nine consist in various more complicated groupings of the Ys ; these nine further combinations exhaust the independent comparisons that can be made, once the first 10 have been laid down as above. There is, however, an infinity of ways of dividing up a group of 20 observations, each yielding a set of 19 independent comparisons between members of the group. The particular arrangement selected depends on the objects of the investigation and not all 19 may need to be isolated. The use to which the isolation of individual degrees of freedom is put will be clear as we proceed with an examination of statistical techniques, but a simple example is not out of place here.

In a group of four observations sums of squares corresponding with the three degrees of freedom can be isolated in the following ways. Twelve are of the form:

$$\tfrac{1}{2}(Y_1-Y_2)^2;\ \ \tfrac{1}{6}(Y_1+Y_2-2Y_3)^2;\ \ \tfrac{1}{12}(Y_1+Y_2+Y_3-3Y_4)^2$$

A further three take the form:

$$\tfrac{1}{2}(Y_1-Y_2)^2;\ \ \tfrac{1}{2}(Y_3-Y_4)^2;\ \ \tfrac{1}{4}(Y_1+Y_2-Y_3-Y_4)^2$$

Another is:

$$\tfrac{1}{4}(Y_1+Y_2-Y_3-Y_4)^2;\ \tfrac{1}{4}(Y_1-Y_2+Y_3-Y_4)^2;\ \tfrac{1}{4}(Y_1-Y_2-Y_3+Y_4)^2$$

Each of these adds up to $SY^2-n\bar{Y}^2$. Those of the first set compare first two Ys, then three, and finally all four. Those of the second set take the observations in pairs, then all together, while the final set involves only simultaneous comparisons of all four together. If we had treated two pairs of animals with two different drugs we could use one of the set of three ways of dividing the results which gives measures of the difference between (Y_1+Y_2) and (Y_3+Y_4), and between Y_1 and Y_2 and then Y_3 and Y_4 separately. If, on the other hand, we had so designed the test that one animal received *both* drugs, one received no drug at all, while each of the remaining two received a single drug, we should need to partition the results as in the last of the three sets above. For, if drug A were given to the animals Y_1 and Y_2, while drug B were given to Y_1 and Y_3, the action of A could be assessed from $(Y_1+Y_2-Y_3-Y_4)$, that of B from $(Y_1+Y_3-Y_2-Y_4)$ and combined effect of A and B from $(Y_1-Y_2-Y_3+Y_4)$, where "combined effect" means those effects not predictable from a knowledge of the actions of A and B when given separately.

In general, an array containing n observations has $n-1$ degrees of freedom. The mean and the variances which have been calculated for the array are called *statistics*, which are estimates of the *parameters*, or constant characters of the population from which the sample has been drawn. The mean is a statistic computed directly from the observations as they stand, and the variance is computed from the second power of these observations. When the variance is computed by subtracting $\bar{Y}T$ from the sum of squares of the items in the array, the factor $\bar{Y}T$ is commonly called the *correction factor* for the origin or mean, or correction for the sum of squares. The term "correction factor for the origin" merely implies that we have taken an arbitrary origin at 0 instead of at the

mean for the purposes of computation. Unless an inordinate amount of labour would be involved, it is usually best to square the Ys directly as recommended and to subtract $\bar{Y}T$ from them, because this avoids the introduction of any possible errors of subtraction. In simple cases there is often little to choose between different methods of computation, but the selection of the best method may be of considerable importance in less simple instances.

CHAPTER 3

COMPARISONS BETWEEN GROUPS

3.1. Differences between groups

In the biological standardisation of insulin, the blood sugar of rabbits is measured. Table 3.1 gives the blood sugars measured in mgm. per 100 ml. of four groups, each containing seven rabbits, the four groups representing four different breeds of rabbit prepared for tests. We wish to know whether these four groups differ in the mean level of blood sugar and thus to determine whether the animals we are going to use for a test form a homogeneous collection.

TABLE 3.1

BLOOD SUGARS OF FOUR DIFFERENT BREEDS OF RABBIT

	Breed							
	1		2		3		4	
	Y_p	$(Y_p-100)^2$	Y_p	$(Y_p-100)^2$	Y_p	$(Y_p-100)^2$	Y_p	$(Y_p-100)^2$
	117	289	137	1,369	135	1,225	109	81
	116	256	136	1,296	122	484	108	64
	128	784	121	441	135	1,225	117	289
	104	16	113	169	138	1,444	118	324
	121	441	145	2,025	131	961	101	1
	100	0	123	529	134	1,156	134	1,156
	123	529	113	169	140	1,600	113	169
Totals, SY_p and $S(Y_p-100)^2$	809	2,315	888	5,998	935	8,095	800	2,084
Means, $\bar{Y}_p$	115·5714		126·8571		133·5714		114·2857	
Correction factor $n_p(\bar{Y}_p-100)^2$		1,697·29		5,049·14		7,889·29		1,428·57
Sum of squares, Sy_p^2		617·71		948·86		205·71		655·43

Total: $2,427·71 = SSy_p^2$

Calling the measurement from an animal in any one group Y_p, we calculate SY_p for each group and from that determine the mean $\bar{Y}_p$ by dividing by $n=7$. The sum of squares for each group is determined by summing $(Y-100)^2$—coding by subtracting 100 from each reading, since no reading is below 100—and from this subtracting the correction factor of $n(\bar{Y}_p-100)^2$ or $(\bar{Y}_p-100)T_p$. The separate variance for each group would be determined by dividing the appropriate sum of squares by $n-1$. However, we are about to test the hypothesis that there is no difference between the groups and for the purposes of the test we assume that the means

and the variances of all four groups are sample estimates of the same statistics. We shall then proceed to test, using the combined data from all groups, whether this assumption can be upheld. Thus we are testing a *null hypothesis*—the hypothesis that the four groups are samples from the same population having the same mean and the same variance.

In order to test this null hypothesis we have available two estimates of variance. The first variance is that of blood sugars within the four groups and is an average of the four separate variances, one of which may be calculated from each group. In order to compute this variance we add the four sums of squares, Sy_p^2, giving us the quantity SSy_p^2, where the symbol SS means "the sum of the sums of." This quantity is 2,427·71. Each of the four groups contributes its own $n-1$ degrees of freedom to the estimate of this variance, and since the groups are equal in number, the total number of degrees of freedom available in estimating the variance is $4(n_p-1)=24$. The variance is therefore $\frac{2,427{\cdot}71}{24}=101{\cdot}2$. This is the combined variance of individual blood sugar readings derived from the distribution of each of the four groups of readings about its own mean, $\bar{Y}_p$.

From this variance we can estimate the variance of the four means. Since each mean is derived from seven observations, the variance of these means will be $\frac{1}{7}$th of the variance of the individual observations, or $\frac{101{\cdot}2}{7}=14{\cdot}4571$.

We can now estimate the variance actually observed between the four means concerned:

$$S(\bar{Y}-{}_p100)^2-\bar{Y}T=257{\cdot}02$$

where $\bar{Y}$ and T are the general mean (122·5714) and the corresponding total. To estimate this observed variance we divide the sum of squares by the three degrees of freedom between the four means, giving us a variance of 85·6733. This variance, estimated from the actual distribution of the means, is thus considerably larger than the variance which we estimate the means to have from the distribution of the individual observations contributing to them.

3.2. The variance ratio

Tables of the function F, known as the *variance ratio*, have been prepared for estimating the significance of differences between two variances. F is equal to the larger variance divided by the smaller

one; in this case it is $\frac{85{\cdot}6733}{14{\cdot}4571}=5{\cdot}93$. Table 3.5, a table of F, gives those values which will be equalled or exceeded in the proportion of cases indicated by the percentage given at the top of the Table with the respective degrees of freedom involved in computing the two variances concerned, if these variances are in fact equal. The number of degrees of freedom used in calculating the greater variance (n_1) is read off along the top of the Table, and the number of degrees of freedom used in calculating the smaller variance (n_2) is read off from the extreme left-hand column. In the present example $n_1=3$ and $n_2=24$. From the Table we see that a variance ratio as high as 4·72 would be encountered only once in 100 times, and from this we conclude that a ratio of 5·93 would, if the two variances were in reality the same, be encountered considerably less than once in 100 times. We must therefore abandon the hypothesis that the observed variance between group means is equal to the variance calculated from the population of individual blood sugars. Hence we conclude that there is a highly significant difference between the mean blood sugars in the different groups of rabbits.

In the interests of clarity we have dealt with the problem in terms of the variance of the means. It is easier in computation to use the reverse procedure and not to reduce the individual variances to variances of means, but instead, when calculating the variance between the means of groups, usually called simply the variance *between groups*, to multiply this variance by the number of observations contributing to each group and thus to estimate a variance per observation instead of per mean. The final form of the analysis set out in Table 3.2 illustrates the method. The sum of squares between groups with three degrees of freedom is 1,799·14, giving a mean square of 599·7, while the mean square within groups with 24 degrees of freedom is 101·2 as before. The total sum of squares, which may be estimated as a check on the arithmetic, should be equal to the two separate sums of squares between and within groups, and the degrees of freedom will, of course, total $n-1$, one less than the total number of individual observations. The total sum of squares, which is $SY^2-\bar{Y}T$, is 4,226·86. The sum of the two sums of squares between and within groups is 4,226·85, and thus the calculations are shown to be correct. The unit discrepancy in the last place of decimals is due to our having rounded off figures in calculation.

TABLE 3.2

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 3.1

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square
Between groups	$n_p S\bar{y}_p^2$	3	1,799·14	599·7
Within groups (error)	SSy_p^2	24	2,427·71	101·2
Total	Sy^2	27	4,226·85	—

$F=5{\cdot}93$; $P<0{\cdot}01$ (1% point is 4·72)

In the example above we have made use of the following algebraical identity:

$$S(Y-\bar{Y})^2=Sn_p(\bar{Y}_p-\bar{Y})^2+SS(Y_p-\bar{Y}_p)^2$$
$$\text{or } Sy^2=Sn_p\bar{y}_p^2+SSy_p^2$$

where Sy^2 is the sum of squares of all deviations from their mean, n_p is the number of observations in group p, and $\bar{Y}_p$ the mean of these observations, y_p the deviation of any one of these observations from $\bar{Y}_p$, and $\bar{y}_p$ the deviation of any one mean from $\bar{Y}$; SSy_p^2 is the sum of the sums of squares of deviations of Y_p from $\bar{Y}_p$. In the example, n_p was the same for all groups and we were thus able to illustrate the calculation of the variance ratio by considering the variance of means or of individual observations in a simple manner. In the next example, n_p is different from group to group, and the advantage of calculating and expressing the results as in Table 3.2, in which the variances are expressed per item, will be apparent.

3.3. The analysis of variance

Table 3.2 is a simple example of the *analysis of variance*, a statistical weapon developed primarily by R. A. Fisher (cf. *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh). The sum of squares of all items from their general mean has been broken down into two parts: the first part expresses the variation encountered between groups or arrays, and the second part that within groups or arrays. In such an analysis we proceed always on the *null hypothesis* and test the assumption of homogeneity by means of the variance ratio. The analysis of variance is a modern statistical procedure which is still being actively developed by professional statisticians and is a powerful yet simple procedure for the analysis of properly designed experiments. Much of the rest of this book will deal with experimental designs based on the fact that the analysis of variance will be used in examining results, for although variance analysis can sometimes be applied to results as presented

by the experimenter to his statistical colleagues, its full application in the reduction of error and the accurate estimation of potencies is only possible when a properly balanced and adequately designed experimental procedure has been followed. Then, in the hands of a competent statistician, computation and analysis are simple and rapid, and the precision of tests will usually be considerably heightened. When only two groups are compared in the analysis of variance, there is only one degree of freedom between the two, and F is thus equal to t^2, where t is the function introduced by "Student" for dealing with the means of small samples. For our purpose t may be completely replaced by F, which deals also with comparisons between more than two groups.

3.4. The analysis of variance with unequal groups

Table 3.3 illustrates computational procedure in the case where groups are unequal. The data in this Table are the weights of young male mice in gm. six weeks after birth. Each group represents one litter in which a large number of males occurred. We wish to know whether the average litter weight differs significantly at this age from litter to litter. No coding is necessary in this example and $T_p = S Y_p$ is determined for each litter and $\bar{Y}_p$ calculated by dividing each total by the corresponding number of animals. The sum of squares within each litter may be calculated by subtracting $\bar{Y}_p T_p$ from $S Y_p^2$ as before, and these calculations are included in Table 3.3 for the purpose of illustration and as a check on subsequent arithmetic. In actual practice, however, the quickest way to perform the computation is to take the total, $T = 290{\cdot}8$, square this total and divide by the total number of mice, giving us $\frac{T^2}{n} = \frac{290{\cdot}8^2}{25}$, which is the same as $\bar{Y}T$. The sum of the squares of the weights of all mice is calculated also, and from it is subtracted $\bar{Y}T$, giving us $Sy^2 = 223{\cdot}10$. This is the total sum of squares. The sum of squares between groups $= S n_p \bar{Y}_p^2$ is also equal to $S \bar{Y}_p T_p - \bar{Y}T$, where $\bar{Y}_p$ and T_p are the means and totals for each group separately.

This sum of squares between groups is 151·41 and the difference between it and the total sum of squares must be equal to the sum of squares within groups. We check this by adding the sums of squares within groups given in the Table, which gives us a figure of $71{\cdot}69 = 223{\cdot}10 - 151{\cdot}41$. Where easy it is useful to perform this actual check, but in many cases such checking is difficult or im-

TABLE 3.3

WEIGHTS OF FIVE LITTERS OF YOUNG MALE MICE IN GM.

Litter:	1	2	3	4	5
	15·0	10·9	10·3	9·2	13·5
	13·4	12·8	10·1	6·7	12·7
	12·7	8·3	8·8	8·9	16·4
	19·2	14·4	11·5	11·0	
	14·3		10·3	10·2	
	14·8			7·6	
				7·8	
$T=SY_p$	89·4	46·4	51·0	61·4	42·6
$\bar{Y}_p$	14·90	11·60	10·20	8·77	14·20
SY_p^2	1,358·02	558·90	523·88	552·38	612·50
$\bar{Y}_pT_p$	1,332·06	538·24	520·20	538·57	604·92
Difference, Sy_p^2	25·96	20·66	3·68	13·81	7·58

$T=290·8$ $SY^2=3,605·68$ $S\bar{Y}_pT_p=3,533·99$ $Sy_p^2=71·69$

$\bar{Y}=11·6320$ $\bar{Y}T=3,382·58$ $\bar{Y}T=3,382·58$ $Sn_p\bar{y}_p^2=151·41$

Sum or difference, $Sy^2=223·10$ $Sn_p\bar{y}_p^2=151·41$ $Sy^2=223·10$

possible, and in the more involved examples with which we shall deal later other means of checking must be employed. From these data we may now construct the analysis of variance given in Table 3.4. There are four degrees of freedom between groups and 20 within groups. Hence the mean square between groups is 37·9, that within groups is 3·58. $F=10·6$ and P, the probability of equalling or exceeding F by chance if there is in fact no difference between the average litter weights, is very small indeed. The 1% point for F is 4·43, with four and 20 degrees of freedom for the respective mean squares, and thus P must be very considerably less than 0·01. We have thus established with a very high degree of significance that the differences between the average weight of male litter mates in these five litters are due not to chance, but to the fact that the members of the different litters are not members of the same "population" of mice.

TABLE 3.4

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 3.3

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square
Between groups	$Sn_p\bar{y}_p^2$	4	151·41	37·9
Within groups	SSy_p^2	20	71·69	3·58
Total	Sy^2	24	223·10	—

$F=10·6$; $P<0·01$

TABLE 3.5

VARIANCE RATIO

20 Per Cent. Points of e^{2z} (H. W. Norton)

n_1/n_2	1	2	3	4	5	6	8	12	24	∞
1	9·47	12·00	13·06	13·73	14·01	14·26	14·59	14·90	15·24	15·58
2	3·56	4·00	4·16	4·24	4·28	4·32	4·36	4·40	4·44	4·48
3	2·68	2·89	2·94	2·96	2·97	2·97	2·98	2·98	2·98	2·98
4	2·35	2·47	2·48	2·48	2·48	2·47	2·47	2·46	2·44	2·43
5	2·18	2·26	2·25	2·24	2·23	2·22	2·20	2·18	2·16	2·13
6	2·07	2·13	2·11	2·09	2·08	2·06	2·04	2·02	1·99	1·95
7	2·00	2·04	2·02	1·99	1·97	1·96	1·93	1·91	1·87	1·83
8	1·95	1·98	1·95	1·92	1·90	1·88	1·86	1·83	1·79	1·74
9	1·91	1·94	1·90	1·87	1·85	1·83	1·80	1·76	1·72	1·67
10	1·88	1·90	1·86	1·83	1·80	1·78	1·75	1·72	1·67	1·62
11	1·86	1·87	1·83	1·80	1·77	1·75	1·72	1·68	1·63	1·57
12	1·84	1·85	1·80	1·77	1·74	1·72	1·69	1·65	1·60	1·54
13	1·82	1·83	1·78	1·75	1·72	1·69	1·66	1·62	1·57	1·51
14	1·81	1·81	1·76	1·73	1·70	1·67	1·64	1·60	1·55	1·48
15	1·80	1·79	1·75	1·71	1·68	1·66	1·62	1·58	1·53	1·46
16	1·79	1·78	1·74	1·70	1·67	1·64	1·61	1·56	1·51	1·43
17	1·78	1·77	1·72	1·68	1·65	1·63	1·59	1·55	1·49	1·42
18	1·77	1·76	1·71	1·67	1·64	1·62	1·58	1·53	1·48	1·40
19	1·76	1·75	1·70	1·66	1·63	1·61	1·57	1·52	1·46	1·39
20	1·76	1·75	1·70	1·65	1·62	1·60	1·56	1·51	1·45	1·37
21	1·75	1·74	1·69	1·65	1·61	1·59	1·55	1·50	1·44	1·36
22	1·75	1·73	1·68	1·64	1·61	1·58	1·54	1·49	1·43	1·35
23	1·74	1·73	1·68	1·63	1·60	1·57	1·53	1·49	1·42	1·34
24	1·74	1·72	1·67	1·63	1·59	1·57	1·53	1·48	1·42	1·33
25	1·73	1·72	1·66	1·62	1·59	1·56	1·52	1·47	1·41	1·32
26	1·73	1·71	1·66	1·62	1·58	1·56	1·52	1·47	1·40	1·31
27	1·73	1·71	1·66	1·61	1·58	1·55	1·51	1·46	1·40	1·30
28	1·72	1·71	1·65	1·61	1·57	1·55	1·51	1·46	1·39	1·30
29	1·72	1·70	1·65	1·60	1·57	1·54	1·50	1·45	1·39	1·29
30	1·72	1·70	1·64	1·60	1·57	1·54	1·50	1·45	1·38	1·28
40	1·70	1·68	1·62	1·57	1·54	1·51	1·47	1·41	1·34	1·24
60	1·68	1·65	1·59	1·55	1·51	1·48	1·44	1·38	1·31	1·18
120	1·66	1·63	1·57	1·52	1·48	1·45	1·41	1·35	1·27	1·12
∞	1·64	1·61	1·55	1·50	1·46	1·43	1·38	1·32	1·23	1·00

Lower 20% points are found by interchange of n_1 and n_2, i.e. n_1 must always correspond with the greater mean square.

Table 3.5 is reprinted from Table V of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

TABLE 3.5—*continued*

VARIANCE RATIO

5 Per Cent. Points of e^{2z}

n_1/n_2	1	2	3	4	5	6	8	12	24	∞
1	161·4	199·5	215·7	224·6	230·2	234·0	238·9	243·9	249·0	254·3
2	18·51	19·00	19·16	19·25	19·30	19·33	19·37	19·41	19·45	19·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·84	8·74	8·64	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·04	5·91	5·77	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·82	4·68	4·53	4·36
6	5·99	5·14	4·76	4·53	4·39	4·28	4·15	4·00	3·84	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·73	3·57	3·41	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·44	3·28	3·12	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·23	3·07	2·90	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·07	2·91	2·74	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	2·95	2·79	2·61	2·40
12	4·75	3·88	3·49	3·26	3·11	3·00	2·85	2·69	2·50	2·30
13	4·67	3·80	3·41	3·18	3·02	2·92	2·77	2·60	2·42	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·70	2·53	2·35	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·64	2·48	2·29	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·59	2·42	2·24	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·55	2·38	2·19	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·51	2·34	2·15	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·48	2·31	2·11	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·45	2·28	2·08	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·42	2·25	2·05	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·40	2·23	2·03	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·38	2·20	2·00	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·36	2·18	1·98	1·73
25	4·24	3·38	2·99	2·76	2·60	2·49	2·34	2·16	1·96	1·71
26	4·22	3·37	2·98	2·74	2·59	2·47	2·32	2·15	1·95	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·30	2·13	1·93	1·67
28	4·20	3·34	2·95	2·71	2·56	2·44	2·29	2·12	1·91	1·65
29	4·18	3·33	2·93	2·70	2·54	2·43	2·28	2·10	1·90	1·64
30	4·17	3·32	2·92	2·69	2·53	2·42	2·27	2·09	1·89	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·18	2·00	1·79	1·51
60	4·00	3·15	2·76	2·52	2·37	2·25	2·10	1·92	1·70	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·02	1·83	1·61	1·25
∞	3·84	2·99	2·60	2·37	2·21	2·09	1·94	1·75	1·52	1·00

Lower 5% points are found by interchange of n_1 and n_2, i.e. n_1 must always correspond with the greater mean square.

Table 3.5 is reprinted from Table V of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

TABLE 3.5—*continued*

VARIANCE RATIO

1 Per Cent. Points of e^{2z}

n_1/n_2	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98·49	99·00	99·17	99·25	99·30	99·33	99·36	99·42	99·46	99·50
3	34·12	30·81	29·46	28·71	28·24	27·91	27·49	27·05	26·60	26·12
4	21·20	18·00	16·69	15·98	15·52	15·21	14·80	14·37	13·93	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·27	9·89	9·47	9·02
6	13·74	10·92	9·78	9·15	8·75	8·47	8·10	7·72	7·31	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·84	6·47	6·07	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·03	5·67	5·28	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·47	5·11	4·73	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·06	4·71	4·33	3·91
11	9·65	7·20	6·22	5·67	5·32	5·07	4·74	4·40	4·02	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·50	4·16	3·78	3·36
13	9·07	6·70	5·74	5·20	4·86	4·62	4·30	3·96	3·59	3·16
14	8·86	6·51	5·56	5·03	4·69	4·46	4·14	3·80	3·43	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·00	3·67	3·29	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	3·89	3·55	3·18	2·75
17	8·40	6·11	5·18	4·67	4·34	4·10	3·79	3·45	3·08	2·65
18	8·28	6·01	5·09	4·58	4·25	4·01	3·71	3·37	3·00	2·57
19	8·18	5·93	5·01	4·50	4·17	3·94	3·63	3·30	2·92	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·56	3·23	2·86	2·42
21	8·02	5·78	4·87	4·37	4·04	3·81	3·51	3·17	2·80	2·36
22	7·94	5·72	4·82	4·31	3·99	3·76	3·45	3·12	2·75	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·41	3·07	2·70	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·36	3·03	2·66	2·21
25	7·77	5·57	4·68	4·18	3·86	3·63	3·32	2·99	2·62	2·17
26	7·72	5·53	4·64	4·14	3·82	3·59	3·29	2·96	2·58	2·13
27	7·68	5·49	4·60	4·11	3·78	3·56	3·26	2·93	2·55	2·10
28	7·64	5·45	4·57	4·07	3·75	3·53	3·23	2·90	2·52	2·06
29	7·60	5·42	4·54	4·04	3·73	3·50	3·20	2·87	2·49	2·03
30	7·56	5·39	4·51	4·02	3·70	3·47	3·17	2·84	2·47	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	2·99	2·66	2·29	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·82	2·50	2·12	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·66	2·34	1·95	1·38
∞	6·64	4·60	3·78	3·32	3·02	2·80	2·52	2·18	1·79	1·00

Lower 1% points are found by interchange of n_1 and n_2, i.e. n_1 must always correspond with the greater mean square.

Table 3.5 is reprinted from Table V of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

TABLE 3.5—*continued*

VARIANCE RATIO

0.1 Per Cent. Points of e^2

n_1/n_2	1	2	3	4	5	6	8	12	24	∞
1	405284	500000	540379	562500	576405	585937	598144	610667	623497	636619
2	998·5	999·0	999·2	999·2	999·3	999·3	999·4	999·4	999·5	999·5
3	167.5	148·5	141·1	137·1	134·6	132·8	130·6	128·3	125·9	123·5
4	74·14	61·25	56·18	53·44	51·71	50·53	49·00	47·41	45·77	44·05
5	47·04	36·61	33·20	31·09	29·75	28·84	27·64	26·42	25·14	23·78
6	35·51	27·00	23·70	21·90	20·81	20·03	19·03	17·99	16·89	15·75
7	29·22	21·69	18·77	17·19	16·21	15·52	14·63	13·71	12·73	11·69
8	25·42	18·49	15·83	14·39	13·49	12·86	12·04	11·19	10·30	9·34
9	22·86	16·39	13·90	12·56	11·71	11·13	10·37	9·57	8·72	7·81
10	21·04	14·91	12·55	11·28	10·48	9·92	9·20	8·45	7·64	6·76
11	19·69	13·81	11·56	10·35	9·58	9·05	8·35	7·63	6·85	6·00
12	18·64	12·97	10·80	9·63	8·89	8·38	7·71	7·00	6·25	5·42
13	17·81	12·31	10·21	9·07	8·25	7·86	7·21	6·52	5·78	4·97
14	17·14	11·78	9·73	8·62	7·92	7·43	6·80	6·13	5·41	4·60
15	16·59	11·34	9·34	8·25	7·57	7·09	6·47	5·81	5·10	4·31
16	16·12	10·97	9·00	7·94	7·27	6·81	6·19	5·55	4·85	4·06
17	15·72	10·66	8·73	7·68	7·02	6·56	5·96	5·32	4·63	3·85
18	15·38	10·39	8·49	7·46	6·81	6·35	5·75	5·13	4·45	3·67
19	15·08	10·16	8·28	7·26	6·61	6·18	5·59	4·97	4·29	3·52
20	14·82	9·95	8·10	7·10	6·46	6·02	5·44	4·82	4·15	3·38
21	14·59	9·77	7·94	6·95	6·32	5·88	5·31	4·70	4·03	3·26
22	14·38	9·61	7·80	6·81	6·19	5·76	5·19	4·58	3·92	3·15
23	14·19	9·47	7·67	6·69	6·08	6·65	5·09	4·48	3·82	3·05
24	14·03	9·34	7·55	6·59	5·98	5·55	4·99	4·39	3·74	2·97
25	13·88	9·22	7·45	6·49	5·88	5·46	4·91	4·31	3·66	2·89
26	13·74	9·12	7·36	6·41	5·80	5·38	4·83	4·24	3·59	2·82
27	13·61	9·02	7·27	6·33	5·73	5·31	4·76	4·17	3·52	2·75
28	13·50	8·93	7·19	6·25	5·66	5·24	4·69	4·11	3·46	2·70
29	13·39	8·85	7·12	6·19	5·59	5·18	4·64	4·05	3·41	2·64
30	13·29	8·77	7·05	6·12	5·53	5·12	4·58	4·00	3·36	2·59
40	12·61	8·25	6·60	5·70	5·13	4·73	4·21	3·64	3·01	2·23
60	11·97	7·76	6·17	5·31	4·76	4·37	3·87	3·31	2·69	1·90
120	11·38	7·31	5·79	4·95	4·42	4·04	3·55	3·02	2·40	1·56
∞	10·83	6·91	5·42	4·62	4·10	3·74	3·27	2·74	2·13	1·00

Lower 0·1% points are found by interchange of n_1 and n_2, i.e. n_1 must always correspond with the greater mean square.

Table 3.5 is reprinted from Table V of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

CHAPTER 4

DOSE-RESPONSE LINES

4.1. Linear regression

In the preceding chapters we have seen how to test for significant differences between the means of arrays. We shall now pass to a consideration of how to relate such differences to other factors which are believed to cause or to be associated with them. We may have established that there are significant differences between the weights of rats of various age groups. How are we to relate age, known to be associated with these differences in weight, to the differences themselves? The statistical methods applicable to such problems are those of *regression*. The regression line is one which relates the changes in one variable to accompanying changes in a second variable. Thus, were we to describe the changes in weight which take place in rats as their age advances by means of a mathematical equation or graphically, we should be dealing with the regression of weight on age.

Regression lines may be straight or curved. We shall confine our attention almost entirely to straight regression lines, i.e. those exhibiting *linear regression*, since it is our endeavour always to deal with the results of biological assay in such a way that the relation of one variable to another may be expressed in linear form. A straight line may be expressed by the equation:

$$Y=a+bX$$

In this regression line, Y is called the *dependent variate* and X the *independent variate*, since the line describes changes which occur in Y as a result of alterations in X. The numbers a and b are constants of the equation, a equalling the numerical value of Y when X is zero, and b, the slope of the line, or the *regression coefficient*, tells us how Y increases (or if b is negative, how Y decreases) as X increases. In Figure 4.1 these relationships are shown in graphical form. It will be seen that a measures the distance up the Y axis, or ordinate, at which the regression line cuts the axis, while b, mathematically speaking, is the tangent of the angle which the line makes with the X axis, or abscissa. In

biological assay X stands for the dose of the drug and Y for the response. For reasons which will be discussed more fully later, X will often be expressed in terms of the logarithm of the dose, i.e. the logarithm of the number of mgm. or ml. of a substance

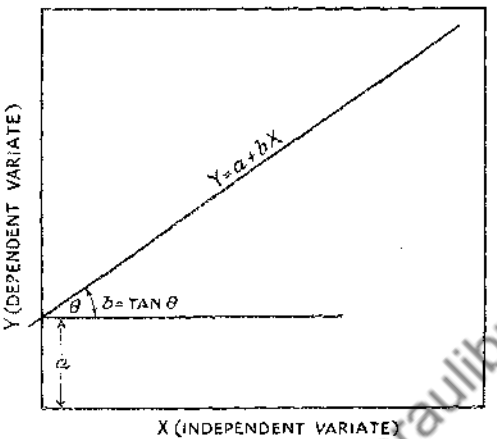


FIG. 4.1. Illustration of linear regression.

under examination, and Y, the response, may be measured in different assays in a great variety of ways.

4.2. The regression of response on dose

In Table 4.1 are shown the results of a test in which five groups of capons, each group containing five birds, were injected with

TABLE 4.1

DOSE-RESPONSE DATA FOR THE INJECTION OF ANDROSTERONE TO CAPONS

	Group 1	2	3	4	5	
Dose in mg. ($X_p=\bar{X}_p$)	0·5	1·0	2·0	4·0	8·0	
Reponses, Y	8	5	13	17	17	
	1	6	7	14	17	
	1	9	12	14	20	
	3	7	10	19	18	
	1	4	11	13	15	
Totals, T_p	14	31	53	77	87	Sum, $T=262$
Means, $\bar{Y}_p$	2·8	6·2	10·6	15·4	17·4	$\bar{Y}=10·48$
SY_p^2	76	207	583	1,211	1,527	$SSY_p^2=3,604·0$
$T_p\bar{Y}_p$	39·2	192·2	561·8	1,185·8	1,513·8	$T\bar{Y}=2,745·76$
Difference, Sy_p^2	36·8	14·8	21·2	25·2	13·2	$Sy^2=858·24$

international standard androsterone (Greenwood, Blyth and Callow, *Biochem. J.*, **29**, 1400, 1935). Androsterone is a pure chemical substance which has the actions typical of the hormone produced by the testis, and responsible for the secondary sexual characteristics of the males of vertebrates. The capon is a castrated cockerel and has only the vestige of comb and wattles. When androsterone is injected into the bird it stimulates growth of the comb and wattles, and these may therefore be used as test objects for the assay of male hormones. In this test the increase in size of the comb was used as an index and expressed as mm. increase in length plus height.

Using the methods of the preceding chapter, we sum the responses and the squares of the responses in each group, giving us a series of totals, T_p, and means, $\bar{Y}_p$, and sums of squares of observations, SY_p^2. The grand total, T, the corresponding mean, $\bar{Y}$, and the corresponding total sum of squares of observations, SSY_p^2, are also written down. Only the totals for sums of squares of observations are given in the Table. The sums of deviations from each group mean and from the general mean, Sy_p^2 and Sy^2, are determined by subtracting, from the corresponding means of squares of observations, the total for each group multiplied by the mean of each group.

Although it is not necessary for the purposes of calculation to determine the individual sums of squares of deviations for each separate group, this has been done to illustrate the *addition theorem* of the analysis of variance and the fact that Sy^2 may be broken down into a series of sums of squares, each associated with a particular source of variation. An analysis of variance showing that the five groups differ significantly in mean response could be built up from the data in Table 4.1, but this will be postponed until further analysis, which follows the calculation of regression, has been made.

We will suppose that we have already tested that the differences between these groups are significant and thus that there has been a meaningful difference in response to the various doses. We now wish to relate mathematically the increase in comb size in the different groups to the dosage of androsterone which they received. We wish further to determine the regression of mean comb size on dose. Readers who are already familiar with the concept of regression will be aware that there are two regression lines which may relate such variables, namely, the regression of Y on X, which we have been considering, and the regression of X on Y. These regressions describe theoretically the average change in Y as X

changes and the average change in X as Y changes. They will not be identical unless the points all fall exactly on a single straight line. In the calculation of regression it is, however, essential that the values of the dependent variate shall not have been arbitrarily selected. If selection has been exercised in the dependent variate, the regression calculated from such material is invalid. Since we choose which doses are to be administered in biological assay, we exercise selection of the dose and it must therefore be used as the independent variate.

The complementary regression, that of dose on response, has been calculated and used in published works through a misunderstanding of the problem. It is understandable that this should have occurred, since in assaying the potency of an unknown preparation we are estimating its activity from the response it has produced, and it would thus seem reasonable to estimate this dose from a regression line which might be thought to tell us the most likely dose corresponding to the series of responses obtained. However, the use of this inverse relationship is a mistake which may invalidate any conclusions drawn from the data.

4.3. Calculation of the regression line

It will be remembered that the mean of an array is so determined that $Sy=0$. It is easy to show that Sy^2 is less when the deviations of individual items are taken from the mean than when they are taken from any other point. When considering a unique set of observations, the two statistics by which we describe this set are so related that the sum of all the deviations from one statistic, the mean, is zero, and the sum of the squares of these deviations is a minimum. This sum of squares divided by the number of degrees of freedom gives us the second statistic, the variance.

When a set of arrays are to be related by a regression line, we proceed by analogy with the above. We determine a line such that the sum of the deviations of all observations from the line is zero. If there is an equal number in each array the sum of the deviations of the means of the arrays from the line is also zero, and the sum of the squares of these deviations is minimal. This is the classical method of *least squares*. We cannot determine the regression line by a simple procedure exactly equivalent to that by which we determine the mean of a set of observations. In calculating the equation of the line we have to employ the fact that Sy^2 is minimal when Sy is zero. When we have calculated the line we have a new

statistic, b, the regression coefficient, which relates changes in one variable to those in the other, and the two lines which may be calculated thus satisfy the two possible criteria, that the deviations of X shall be fitted by least squares, and that the deviations of Y shall be fitted by least squares. It has just been explained that we are interested only in the computation of the second of these two regressions.

A regression line calculated according to the principle of least squares passes through the point $\bar{X}$, $\bar{Y}$. If E is the value which we predict from the line should be the response to a given dose, the regression equation which we have fitted is

$$E=a+bX$$

Since the point $\bar{X}$, $\bar{Y}$ lies on the regression line, E at this point is equal to $\bar{Y}$. Hence

$$\bar{Y}=a+b\bar{X} \text{ and } a=\bar{Y}-b\bar{X}$$

Substituting this value of a in the equation we see that

$$E=\bar{Y}-b\bar{X}+bX \text{ or } E-\bar{Y}=b(X-\bar{X})$$

from which we have eliminated a. From the principle of least squares it may also be shown that the best estimate of b is given by the equation

$$b=\frac{S(X-\bar{X})(Y-\bar{Y})}{S(X-\bar{X})^2}=\frac{Sxy}{Sx^2}$$

If n_p is constant,

$$b=\frac{S(\bar{X}_p-\bar{X})(\bar{Y}_p-\bar{Y})}{S(\bar{X}_p-\bar{X})^2}=\frac{S\bar{x}_p\bar{y}_p}{S\bar{x}_p{}^2}$$

With these equations we can calculate the regression directly from the observations. Each dose in mgm. is converted to its corresponding logarithm, using the most convenient base for the computation in hand. If the doses are irregularly spaced, there is no point in using any other than base 10 and the logarithms are read off from an ordinary table. If, as in this case, the doses are so spaced that their logarithms increase in arithmetical progression, then the most convenient base to use is that given by the ratio of two successive doses. Each successive dose in Table 4.1 is twice the dose preceding it and we therefore take logarithms to base 2 and call these the *coefficients* representing the log dose. These are the successive values of $\bar{X}_p$, which, since all doses in a group are the same, equals X_p. In order to calculate the regression line we

determine $S\bar{x}_p^2$ and $S\bar{x}_p\bar{y}_p$, the sums of squares of deviations of doses from the mean dose, and of the products of the corresponding deviations of doses and responses from their respective means. We saw earlier that

$$Sx^2 = SX^2 - n\bar{X}^2$$

The analogous identity relating sums of products is

$$Sxy = SXY - n\bar{X}\bar{Y}$$

where n is the total number of observations. The mean product of xy, which is $\frac{Sxy}{n-1}$, analogous to the variance of a single variable, is called the *covariance*, and b, the regression coefficient, is the covariance divided by the variance of X. Since it is unnecessary directly to determine the covariance and variance by dividing the sums of squares and products by the common factor $n-1$, we use the equation derived above, so that for Table 4.2, $b = \frac{38{\cdot}4}{10}$, or 3·84. Substituting in the equation

$$E - \bar{y} = b(X - \bar{X})$$

we find that

$$E - 10{\cdot}48 = 3{\cdot}84(X - 1)$$

whence

$$E = 6{\cdot}64 + 3{\cdot}84X$$

From this equation we may calculate the value of E for each dose, and this is tabulated under the estimated response in Table 4.2.

TABLE 4.2

CALCULATION OF REGRESSION WITH CHECK FOR TABLE 4.1

Dose	Coefficient for log dose $\bar{X}_p = \log_2$ (dose)	Response, $\bar{Y}_p$	$\bar{X}_p^2$	$\bar{X}_p\bar{Y}_p$	Estimated response, E	$n_p(\bar{Y}_p - E)^2$
0·5	−1	2·8	1	−2·8	2·80	0·000
1·0	0	6·2	0	0·0	6·64	0·968
2·0	1	10·6	1	10·6	10·48	0·072
4·0	2	15·4	4	30·8	14·32	5·832
8·0	3	17·4	9	52·2	18·16	2·888
Totals	5	52·4	15	90·8	52·4	9·760
Means	1	10·48	—	—		
Correction for mean			5	52·4		
Sums of squares and products			10	38·4		

$b = \frac{S\bar{x}_p\bar{y}_p}{S\bar{x}_p^2} = \frac{38{\cdot}4}{10} = 3{\cdot}84.$

The square of the error of each estimate $(\bar{Y}_p - E)^2$ is listed in the next column as a check on the calculations. The sum of the errors of estimate is zero, i.e.

$$SE = S\bar{Y}_p$$

and the sum of the squares of the errors of estimate, $Sn_p(\bar{Y}_p - E)^2$, which from the definition of the method is lower than that which would be given by any other straight line, is 9·760. Note that this sum of squares is computed with regard for the number in each group. It is unnecessary, except for the purpose of interest and for checking, to calculate the sum of the squares of the errors of estimate by this method, since they may be calculated directly from the relationship

$$Sn_p(\bar{Y}_p - E)^2 = Sn_p\bar{y}_p^2 - \frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p^2}$$

$$= Sn_p\bar{y}_p^2 - \frac{n_p(S\bar{x}_p\bar{y}_p)^2}{S\bar{x}_p^2},$$

since n_p is constant.

This quantity comes to 9·760, as in Table 4.2.

4.4. The analysis of variance, including regression

We are now in a position to tabulate the analysis of variance for these data. We have split the sum of squares into three parts. These are listed in Table 4.3.

TABLE 4.3

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 4.1

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square	F	P
A. Linear regression	$\frac{n_p(S\bar{x}_p\bar{y}_p)^2}{S\bar{x}_p^2}$	1	737·30	737·3	132·6	<0·001
B. Deviations from regression	$Sn_p\bar{y}_p^2 - A$	3	9·76	3·25	0·58	>0·05
C. Random sampling (error)	SSy_p^2	20	111·20	5·56	—	—
Total	Sy^2	24	858·26	—	—	—

Part A is that which can be accounted for by the regression of response on dose and with it is associated one degree of freedom—that "used up" in calculating the regression coefficient.

Part B represents the deviations from this regression line which the means $(\bar{Y}_p)$ exhibit. With these deviations are associated three

degrees of freedom, two less than the number of groups. The total number of degrees of freedom between groups is four, one of which has already been accounted for in the calculation of the regression line.

Part C, called random sampling, is the total sum of squares within groups (SSy_p^2) and with it are associated the 20 degrees of freedom, four within each group. These three sums of squares should and do add up to the total sum of squares (Sy^2) and the number of degrees of freedom add up to the total number of degrees of freedom in the whole set of observations, or 24 degrees of freedom. The mean square or variance associated with A, B and C is obtained by dividing by the appropriate number of degrees of freedom and is tabulated in Table 4.3. The first two mean squares are to be compared with the third, due to random sampling, which we use as our estimate of the error inherent in the determination of the statistics computed. F, the variance ratio, is obtained by dividing these by the mean square 5·56 and takes the values 132·6 and 0·58 respectively. The first value is very large and P, the probability of equalling or exceeding it if there were in fact no real regression of response on dose, is infinitesimal. The second value of F is less than unity, which means that the deviations from regression are smaller than would be expected from the degree of variation shown within groups. Values of F less than unity will not be found in the Tables of F, and in order to test the significance we take the reciprocal of the value, which is 1·72, and enter the Table of F with n_1 equal to the number of degrees of freedom in the larger mean square, in this case $n_1=20$, and n_2 equal to the number of degrees of freedom associated with the smaller mean square, whence $n_2=3$. We then find that P is greater than 0·05, which means that the deviations from regression, although rather smaller than would be expected, are not so small that any significance is to be attached to the fact. From this analysis we thus conclude that there is a highly significant regression coefficient relating response to dose and that the points we have determined fit a log dose-response line perfectly well. This is illustrated in Figures 4.2 and 4.3.

4.5. Log dose-response line

The log dose-response line had the equation

$$E=6\cdot64+3\cdot84X$$

in which X, the coefficient for log dose, is the logarithm to base 2

of the dose. There are no logarithmic tables computed to base 2, but this need occasion no alarm, since a simple transformation is available by which we may calculate the value of E for any chosen dose, using ordinary tables of logarithms. The relationship in question is

$$\log_{10}X=(\log_N X)(\log_{10}N),$$

where N is any number. Hence, in order to convert the regression equation to one in which the logarithm of the dose is taken to base

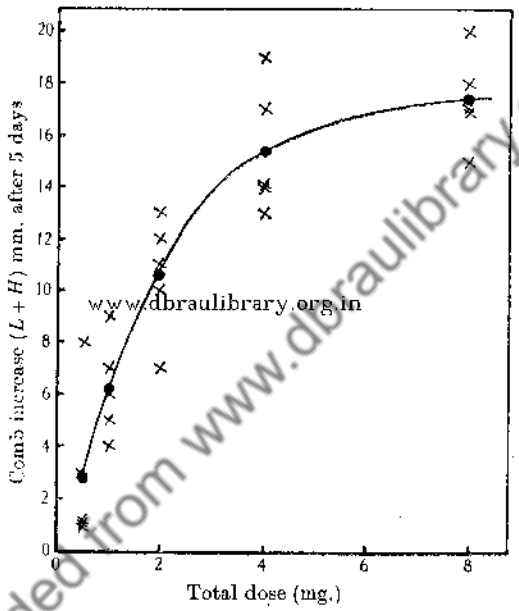


FIG. 4.2. Dose-response curve for the data of Table 4.1. (From Greenwood, Blyth and Callow, *Biochem. J.*, **29**, 1,400, 1935.)

10, we divide the term involving X by $\log_{10}N$. The equation now becomes:

$$E=6{\cdot}64+\frac{3{\cdot}84X}{0{\cdot}3010}=6{\cdot}64+12{\cdot}76X$$

4.6. Regression with unequal groups and spacing of doses

In calculating the above regression we were dealing only with groups containing equal numbers of animals which were given dosages in geometrical progression and computation was relatively simple. It sometimes happens either through bad planning or unavoidable accident that the numbers vary from group to group.

It may also happen, usually through lack of foresight, that the doses are not in geometrical progression and thus do not lend themselves to easy calculation. We shall now consider how to deal with the calculation of regression in a case where the number of observations in each group is not the same and where the doses are not on a convenient scale. In Table 4.4 are shown the results of a

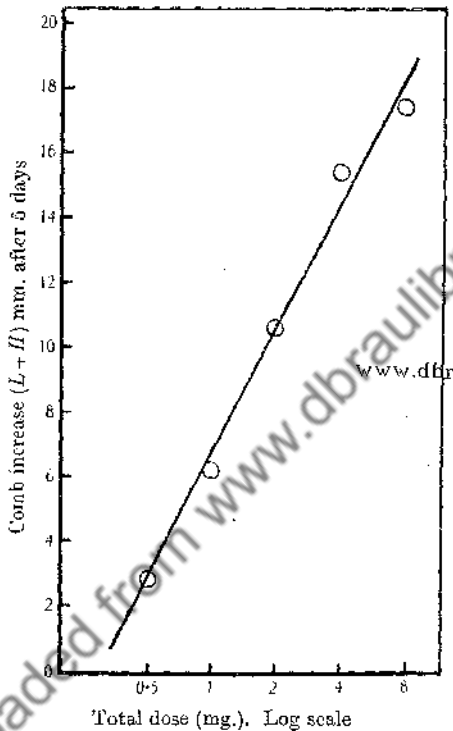


FIG. 4.3. The data of Fig. 4.1 plotted as response against log dose. (From Greenwood, Blyth and Callow, *Biochem. J.*, **29**, 1,400, 1935.)

test in which five different groups of rats, which had been maintained on a diet deficient in vitamin A, were then given supplementary rations of vitamin A for a period of five weeks. The vitamin was given in the form of cod-liver oil and doses are measured in terms of the number of mgm. of cod-liver oil given daily to each group. When little or no vitamin A is included in the diet the animals lose weight. On supplementing their diet with cod-liver oil, rats gained weight in approximate proportion, up to a certain maximum, to the

logarithm of the dose. The data in Table 4.4 are the records of an actual experiment which were published without any information about the individual gains in weight (Coward *et al.*, *Biochem. J.*, **24**, 1,952, 1930). Thus for each group we know only the number of rats, the mean gain or loss in gm. and the dose which rats in the group received. We therefore cannot calculate Sy^2, nor for the purposes of illustrating the calculation of regression is it necessary to do so.

TABLE 4.4

CALCULATION OF REGRESSION OF MEAN WEIGHTS OF VITAMIN-DEFICIENT RATS ON DOSAGE OF COD-LIVER OIL.

Dose in mg.	$\bar{X}_p$ = $\log_{10}$ (dose)	Mean weight in gm., $\bar{Y}_p$	No. in group n_p	$n_p\bar{X}_p$	$n_p\bar{Y}_p$	$n_p\bar{X}_p^2$	$n_p\bar{X}_p\bar{Y}_p$	$n_p\bar{Y}_p^2$
0·25	−0·602	−11·5	31	−18·662	− 356·5	11·235	214·61	4,099·8
1·0	0·000	13·0	37	0·000	481·0	0·000	0·00	6,253·0
1·5	0·176	17·1	35	6·160	598·5	1·084	105·34	10,234·4
2·5	0·398	27·7	32	12·736	886·4	5·069	352·79	24,553·3
7·5	0·875	48·2	31	27·125	1,494·2	23·734	1,307·43	72,020·4
Totals			166	27·359	3,103·6	41·122	1,980·17	117,160·9
Means			—	0·16481	18·6964	—	—	—
Correction for mean						4·509	511·51	58,026·1
Sums of squares and products						36·613	1,468·66	59·134·8

$$b=\frac{Sn_p\bar{x}_p\bar{y}_p}{Sn_p\bar{x}_p^2}=\frac{1,468{\cdot}66}{36{\cdot}613}=40{\cdot}133.$$

The logarithm to base 10 of each dose is noted under $\bar{X}_p$. It is only necessary to write this logarithm to the third place of decimals and it will be seen that the logarithm of the smallest dose, 0·25 mgm., is a negative number, ascertained by subtracting the logarithm of 2·5 from 1 and changing the sign. We do not leave this logarithm in the form $\bar{1}$·398, as we would in using it for common multiplication or division. $\bar{Y}_p$ is then written down against each $\bar{X}_p$. Again it will be noted that the group on the lowest dose lost weight and thus the gain in weight is negative.

The number of animals in each group, n_p, is now tabulated and is to be used as a weighting factor by which to weight the observations in each group in the calculation of regression. Each animal in the test is supposed to contribute unit amount of information, which it will do as long as the variance is independent of the

response; and thus the value of a group in determining the slope and position of the dose-response line is directly proportional to the number of animals it contains. The next two columns, under the headings $n_p\bar{X}_p$ and $n_p\bar{Y}_p$, are filled in by multiplying the log dose and response respectively by the number of animals in each group, and the corresponding weighted means are determined by dividing the sums of these columns by n, the sum of all the n_ps. The last three columns in the Table, which are the sums of squares and products as before, are each weighted by multiplying the individual entries by the corresponding n_p and the totals are recorded at the bottom of the column. In order to obtain the weighted sums of squares and products, we subtract in the case of the sums of squares the totals multiplied by the weighted means ($T\bar{X}$ and $T\bar{Y}$) and n times the product of the two weighted means (this equals either total times the complementary weighted mean in the case of the sum of products). These corrections when subtracted from the totals give us the weighted sums of squares and products. b, the regression coefficient, is then

$$\frac{Sn_p\bar{x}_p\bar{y}_p}{Sn_p\bar{x}_p{}^2}=40{\cdot}133.$$

The dose-response line is

$$E-18{\cdot}696=40{\cdot}133\ (X-0{\cdot}16481)$$

$$E=12{\cdot}082+40{\cdot}133X,$$

and is shown in Figure 4.4. That part of the variance of the means which is attributable to linear regression is

$$\frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p{}^2}$$

which is 58,912·3. The sum of the squares of the deviations of the means from $\bar{Y}$ on a per item basis is:

$$Sn_p\bar{y}_p{}^2=59{,}134{\cdot}8$$

and the difference between these two sums, which is the sum of squares attributable to deviations from regression, is 222·5. The accuracy of these calculations may then be checked as in Table 4.5, where the estimated responses are compared as in Table 4.2 with the actual mean responses. It will be seen that the total errors of estimate in this Table add up to −0·06, the difference from zero being accounted for by the rounding of numbers used in

constructing Table 4.5, and our total for the sum of squares of errors of estimate is 221·75, which agrees well enough in this rough

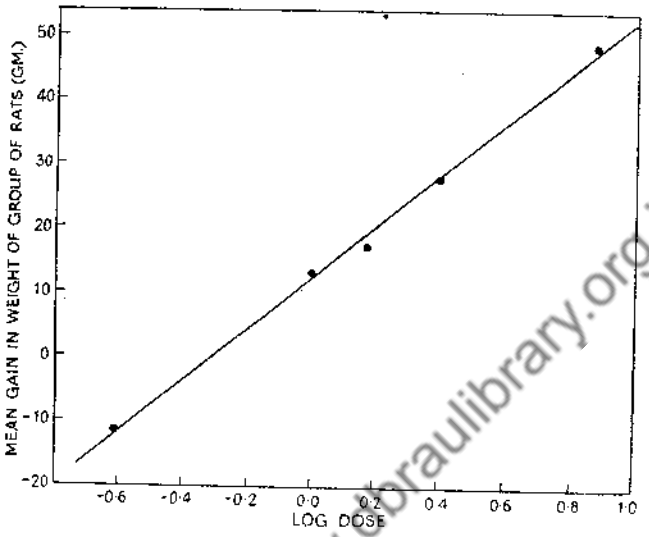


FIG. 4.4. The relation between the mean gain in weight of groups of vitamin-A deficient rats and the log dose of cod-liver oil (Table 4.4).

check with the figure above, but serves also to illustrate that when calculating to a high degree of accuracy, it is desirable to retain more figures after the decimal point than those in Table 4.5.

TABLE 4.5

CHECK ON CALCULATIONS OF TABLE 4.4

Mean response $\bar{Y}_p$	Estimate, E	Error of estimate, $(\bar{Y}_p - E)$	$n_p(\bar{Y}_p - E)^2$
−11·5	−12·06	−0·56	9·72
13·0	12·09	−0·91	30·64
17·1	19·15	2·05	147·09
27·7	28·05	0·35	3·92
48·2	47·19	−0·99	30·38
	Totals	−0·06	221·75

Table 4.6 gives the analysis of variance for these data, analogous to that in Table 4.3, except that there is no item under the heading "random sampling," since we do not know the values of the

individual observations. When confronted with such material we may estimate the significance of the regression coefficient by using the deviations from regression as error. We do not know whether these deviations are significant, but if they are it means merely that we are using a larger mean square against which to test the significance of the mean square attributable to linear regression than might be the case if we were fully informed as to error due to random sampling. Indeed, if these deviations are significant and in addition are not systematic, and give no indication that the data depart in some regular way from a straight line, lying for instance on a smooth curve, the mean square associated with deviations from regression should be used in assessing the significance of the regression coefficient, even if an error mean square is available. Under these circumstances we are faced with significant heterogeneity of the test material, and the situation is not satisfactory. Only if the deviations could be found to be significantly less than those expected on a random sampling basis would their use as the error term prove to have been fallacious. This situation would, however, be unlikely in practice and the use of deviations from regression in assaying the significance of the regression is quite justifiable when no other information is available. The value of F in Table 4.6 is very high (792·9) and we thus have every confidence in the extremely high significance of the regression coefficient (it exceeds the 0·1% level of significance for $n_1=1$; $n_2=3$).

TABLE 4.6

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 4.4

Source of variation	Formula	Degrees of freedom	Sum of Squares	Mean square	F	P
A. Linear regression	$\frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p^2}$	1	58,912·3	58,912·3	729·9	<0·001
B. Deviations from regression (used as error)	$Sn_p\bar{y}_p^2-A$	3	222·5	74·3	—	—
Total	$Sn_p\bar{y}_p^2$	4	59,134·8	—	—	—

CHAPTER 5

FURTHER DISCUSSION OF DOSE-RESPONSE LINES

5.1. The log dose-response relationship

When the response to a series of graded doses of a pharmacologically active substance is plotted against the doses on an arithmetic scale, the line relating these two variables is usually curved. Very low doses may elicit no or practically no response, and high doses may elicit responses which differ very little from one another if the maximum response which the test object is capable of exhibiting has been reached. Between these two levels

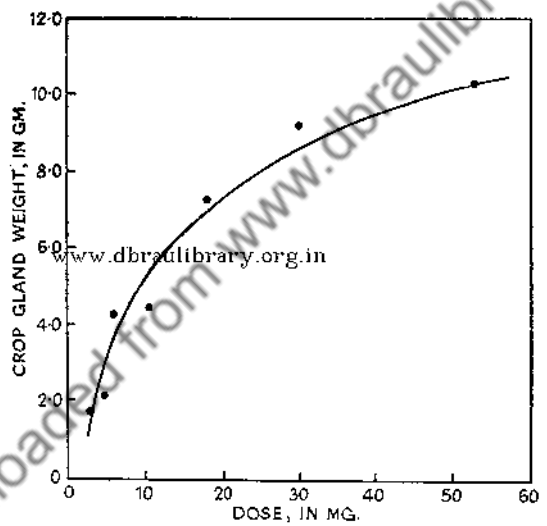


FIG. 5.1. The response of groups of pigeons to various doses of prolactin. Prolactin stimulates the growth of the crop-gland, the weight of which is used in assaying preparations of the hormone. (From Emmens, *J. Endocrinol*, **2**, 194, 1940.)

will be a range over which changes in response as the dose is increased occur most rapidly, and it is this portion of the curve which is used for the purpose of biological assay. Sometimes, as far as can be determined, very small doses take effect and there may be no apparent region of a curve over which they are ineffective. Some examples of typical dose-response curves are shown in Figures 5.1–5.3. The curves fitted to the points are logarithmic,

and do not pass through the "response" to zero dose. It is characteristic of the middle ranges of such dose-response curves,

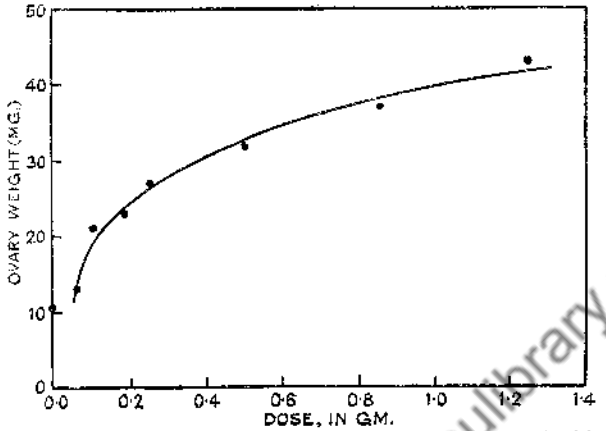


FIG. 5.2. The mean ovary weight of groups of female rats injected with a gonadotrophic preparation of pregnant women's serum. (From Emmens, *J. Endocrinol*, **2**, 194, 1940.)

where the curve is steepest, that equal increments in response are not produced by equal increments in dose, but only by equal

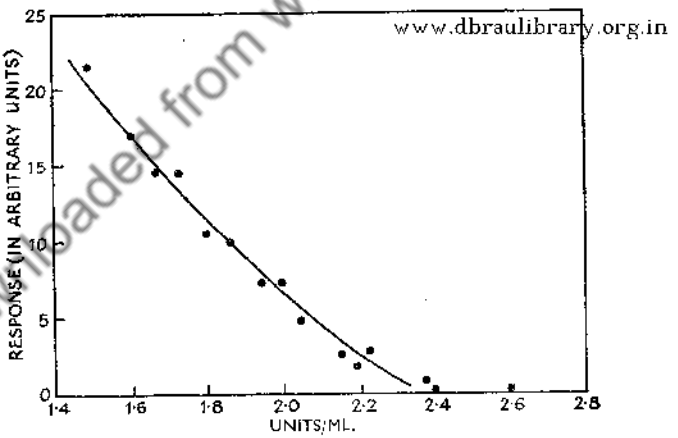


FIG. 5.3. The relation between the growth inhibition produced in plate-droplet cultures of *Staph. aureus* and the dose of penicillin in international units. The response is measured in grades approximately corresponding to the width of a ring of growing *staphylococci*. The growth of this ring is smaller, the higher the dose of penicillin.

multiples of the dose. Thus, while response is increasing in approximately arithmetic progression, the corresponding doses are

increasing in approximately geometric progression. This has long been recognised as characteristic of a large number of pharmacologically active substances, which are then said to follow the Weber-Fechner law, so called after its first discoverers. Dose-response curves which follow this law may be converted to straight lines by the simple transformation of dose to log dose. In plotting the curves the logarithm of the dose may be read from tables and plotted along the abscissa instead of the dose itself, or specially prepared logarithmic paper may be used to simplify the making of the diagram. In such paper one or more cycles to base 10 are plotted logarithmically on the sheet, so that it is unnecessary for graphical purposes to determine the actual logarithms of the doses.

The statistical handling of dose-response data has grown up around the almost universal log dose-response relationship. It so happens that this naturally observed relationship lends itself to ease in computation, particularly in the determination of relative potencies. Occasionally, dose-response data deviate so much from a linear log dose-response relationship that the latter cannot be used for computation. Such cases often present difficulties for the statistician, and the use of other formulae will most frequently lead only to approximate methods being available in the computation of relative potencies and of error. Thus, it would often be preferable to fit the observed data to a log dose-response line over, perhaps, a limited range of response and to take into account the deviations which such data may show from a directly linear function in the estimation of error. When the dose in arithmetic units is linearly related to the response, the methods described in Chapter 20 may be applicable.

A little thought will make it quite clear that the log dose-response relationship cannot hold over the whole range of dosage in any instance. There is never an unlimited maximum to which the response may rise, whilst the lower part of the curve, which cuts the X axis at some point greater than $X=0$ and then plunges rapidly down asymptotically to the Y axis, cannot describe any real relationship between dose and response. This does not affect the usefulness of the curve over the region in which it is to be used, and over which it is in the majority of cases an extremely good approximation to whatever may be the true form of the relationship. Figures 5.4–5.6 show the curves of Figures 5.1–5.3 plotted on the logarithmic scale.

Thus, in dealing with the conversion of dose to log dose in

statistical computation, it must be borne in mind that the relationship which we choose to employ may have no higher status than that of an avowedly convenient method of graduation. Biometricians

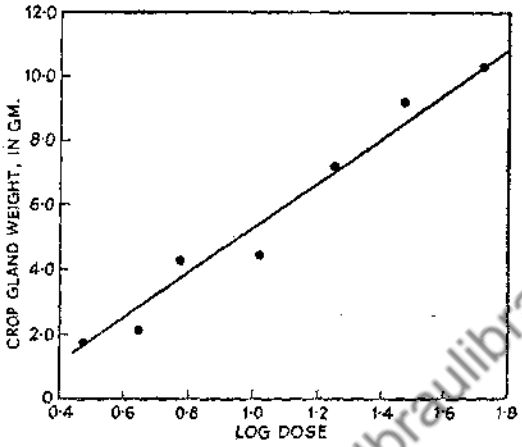


FIG. 5.4. The data in Fig. 5.1 plotted on a log scale.

often choose to express the response in a variety of sometimes exotic ways, and although there may be good reasons for believing that the middle ranges of many dose-response curves in which such

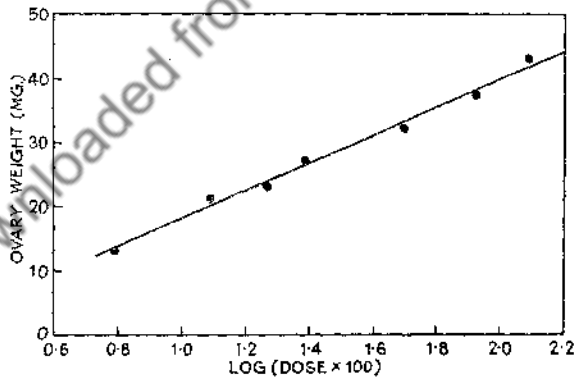


FIG. 5.5. The data in Fig. 5.2 plotted on a log scale.

a simple quantity as the weight of an animal or the amount of gas evolved in a chemical reaction should follow the Weber-Fechner law, there may be no reason at all why some quite artificial index of response should do so. Responses based on a series of arbitrary

grades, an example of which is the grading 1, 2, 3, etc., given to a series of changes which occur as the result of the administration of a drug, may fit a logarithmic curve by chance, but there is no inherent reason why they should do so. The grading may, of course, be modified in the light of the relationship subsequently discovered between dose and response in order that the new relationship, as measured by the grading, may be more suitable.

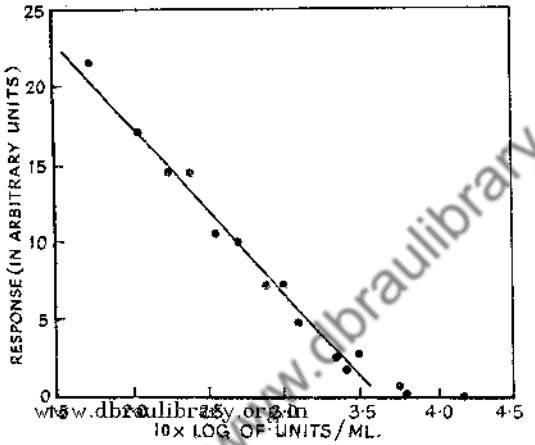


FIG. 5.6. The data in Fig. 5.3 plotted on a log scale.

5.2. Using the relationship in assays

In relating the potency of one drug or preparation of a drug to that of another, we must satisfy ourselves that the actions of the two are identical. Apart from the usual biological control which will be exercised to ensure that such is the case, a properly designed test will include a statistical check on this fact. The check consists in establishing that the same dose-response curve may be fitted to the two separate substances. When dealing with a linear log dose-response relationship, this implies only that the two dose-response lines shall be parallel. If it is found in practice that they are not, it means, quite simply, that under the conditions of the test the one substance cannot be assayed with any accuracy in terms of the other, since our estimate of relative potency will depend on the particular level of response at which it is made. Thus, in Figure 5.8, which illustrates the case in which two log dose-response lines are not parallel, the potency of substance A measured at a 40-mgm. response level would be twice that of substance B, whereas if the

measurement is made at 80 mgm. response the potency of substance A is approximately four times that of substance B. When

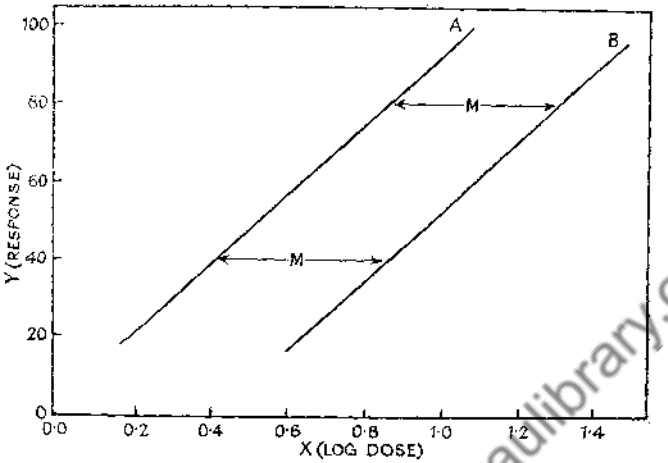


FIG. 5.7. An assay with parallel log dose-response lines. The log ratio of potency, M, is constant at all levels of response of substances A and B.

the two lines are parallel, M, the logarithm of the relative potency of the two substances is the distance between the two lines, measured parallel to the abscissa.

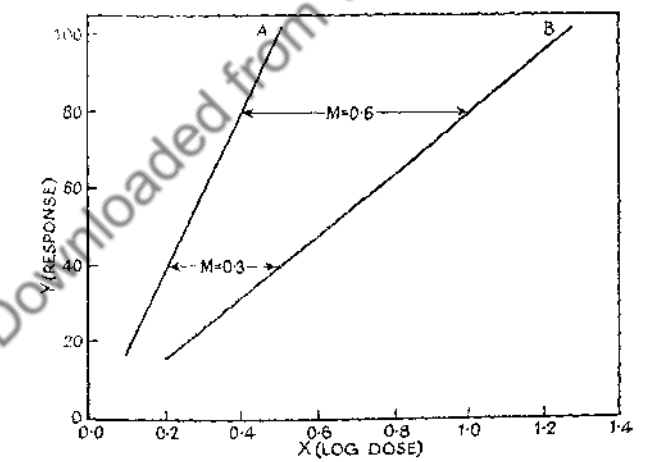


FIG. 5.8. A test with log dose-response lines not parallel. The log ratio of potency, M, depends on the level of response at which it is measured.

When planning the setting-up of tests involving the establishment of a dose-response curve or curves, the following points should be

borne in *mind*. If two groups of animals are given two different doses of a substance, the two points $\bar{X}_1$, $\bar{Y}_1$ and $\bar{X}_2$, $\bar{Y}_2$ may be fitted by a straight line simply by joining them, and there will be no test available of the adequacy of the log dose-response line or any other line, except lines fitted on some assumption involving the "response" of an untreated group. It will be noted that observations from a control group receiving no treatment cannot be fitted to a line relating the logarithm of the dose to response. Such a control group is useful, however, in showing whether the group exhibiting the lowest "response" is in fact responding at all. If it is not, it too is of no use for the purposes of assay.

Three or more treated groups enable us to test the linearity of the relationship, and it is in general advisable to use at least four groups and preferably more in preliminary investigations to decide whether a proposed method of assay can reasonably be based on the hypothesis of linearity of the log dose-response relationship. With three groups, departure from linearity of the form of simple curvature may be detected and with more than three groups the possibility of departures of a more complex type can be examined. We have already seen in the preceding chapter how such tests may be conducted by use of the analysis of variance. We establish first that there is a significant regression of response on dose and in the light of the variance of Y we test whether the errors of estimate, $Sn_p(\bar{Y}_p - E)$, can be accounted for by random sampling or whether they are too large for this to be likely. If they are too large for the log dose-response relationship to be tenable, we may allow for the excess departure in our estimate of error, or we may attempt to modify the way in which we measure or express the response so as to obtain a linear relationship, or as a last resort we may be compelled to abandon the log dose-response relationship altogether.

These are the kind of preliminary investigations which will most usually be made with a substance intended for use as a standard preparation when we are investigating the possibilities of determining potencies by any particular method. When we are satisfied that a linear relationship with reasonably small errors has been established for any particular method of assay, we shall then proceed to set up tests for comparing the potencies of other substances with that of the standard, and while we shall normally include sufficient doses of both unknown and standard at least to test the validity of the assumption that the two are fitted by the same regression line, it may not be necessary to establish complete dose-response lines

for either the standard or unknown in the course of subsequent assays.

Some of the dose-response lines used for the purposes of illustration in this monograph do not have the doses arranged in geometrical progression. They have been taken from the literature for illustrative purposes, and, except in so far as it facilitates computation, it is not essential that in the establishment of original dose-response lines the doses should be arranged in any particular order of magnitude as long as they cover a wide enough range. However, it is not sufficiently realised that the full value of modern statistical treatment can only be attained in biological assay when the doses of both the unknown and the standard are distributed systematically. They must be distributed in geometric progression, which means that their logarithms must be in arithmetic progression. Failure to plan an assay along these simple lines leads not only to a very much more complicated analysis than should have been necessary, but may also make it impossible to obtain as much information from the data as could have been obtained. The great ease of computation to which this arrangement lends itself may be realised when we consider that, by taking our logarithms to a base equal to the multiple by which doses are increasing, we can express the log dose by the natural numbers, 1, 2, 3, 4, etc. Our final estimate of relative potencies will not normally be a cardinal number, and in order to convert this logarithm back to the equivalent dose we have to remember the relationship in Chapter 4. The use of such a conversion factor is frequently well worth while, as the few logarithmic transformations which have to be made at the end of our calculations take little time compared with the increased computation which would have been necessary had logarithms been taken to base 10 or any other base.

CHAPTER 6

RESTRICTIONS IN DESIGN

6.1. Randomisation of test objects

One of the variables in biological assay is always arbitrarily selected. This variable is the dose of the drug which is to be administered to particular groups of test objects. Other variables which may be capable of influencing the course of the test are frequently subject to selection as well. In the description of the majority of biological assays a considerable amount of space is given to defining the conditions under which the test shall be made and describing the class of test object suitable for the test. Animals of only one sex and between certain ages and weights may be prescribed; the test must be conducted within certain limits of time and perhaps at certain temperatures; and the response must be measured subject to other conditions, more or less rigid. Thus, although the question of designing tests in which certain rigidities of this kind may be relaxed will be discussed later, it is generally speaking true that the more rigid these initial conditions are made, the more accurate is the test.

Once these initial specifications have been fulfilled, it is then necessary to see that the allotment of test objects to doses and to any other restricted classes which may be included in the design of the test shall take place by a process called randomisation. Allotment at random is comparable to the process of thorough shuffling and dealing from a pack of cards. In the simplest case, where we wish, for instance, to allot five groups each of 20 animals to five different doses, we could do so by writing the number of each animal on a slip of paper and drawing them one after another from a hat, allotting each animal in turn to group 1, 2, 3, 4 or 5. This is rather a tedious process where large numbers are concerned and to save the experimenter the trouble of carrying out the actions, tables of random numbers have been prepared (as in Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver & Boyd, Edinburgh) by which the test objects may rapidly be allotted to classes.

Many workers have been under the impression that such a procedure as taking the first 20 animals that come to hand from a cage and allotting them to the first dosage group, taking the next 20 and allotting them to the second dosage group and so on, constitutes random selection. This is most definitely not the case. The first 20 animals that come to hand will often be the tamest animals. They may be the biggest animals and they will quite rarely be representative of the group as a whole. A striking instance of this occurred when an assistant was requested to select groups of mice at random, and it was afterwards possible to demonstrate a highly significant correlation between the order in which the animals were taken from the cage and the weight of the animals. Similarly, it is not random selection to allot the top rack in an animal room to one dose, the second rack to another, and so on, because the position of the animals in the room will sometimes affect their response in the tests. The top of the room may be lighter than the bottom; one wall may be warmer than another; and animals in the one position may even receive more food than those in another if assistants feed them in a set routine; and these are factors likely to affect the results of a large number of tests. The order in which doses are administered may also affect results: this is particularly likely to happen when the response is measured within a short time after administration or if the drug is given at a certain period after preparation of the test object. Thus, whenever such factors are even remotely likely to affect results, the order of administration of doses should also be determined by a process of strict randomisation. It should be noted also that attempts to adjust groups of animals so that their mean weights shall be approximately the same are open to criticism. Methods of making such adjustment and of allowing for differences which may be found to exist which are more statistically acceptable will be described later on.

6.2. Allotting test objects to groups

It will be seen that in effect the process of randomisation must fulfil the requirement that any particular test object shall be as likely to be allotted to any particular class as to any other, and conversely that any particular class shall be as likely to receive any particular test object as is any other class. These conditions may, however, be fulfilled within the limits of certain planned restrictions in the design of the test. This important fact is used

in designing tests which shall be so balanced that the effect of a number of variables may simultaneously be examined in subsequent statistical analysis. If, for instance, we are dealing with several litters of rats which are known or believed to exhibit less variation within litters than between litters, we shall normally desire so to distribute litter mates that the smaller variation between members of the same litter may be used in reducing the errors of the test.

In a simple instance, we may have selected litters from each of which four animals may be drawn, and each animal is to be allotted to a different dose. This is perfectly permissible as long as the individual members of litters are allotted *at random* to the four different doses, for it will be seen that the conditions outlined above have not been violated, as each dose is as likely to be administered to any particular animal as any other dose. When the results of the test are being examined, the differences in response between litters may nevertheless be segregated from other sources of variability and eliminated from our estimate of error. It would be an entirely mistaken procedure to allot the whole of one litter to one dose and other whole litters to the other doses, on the assumption that since the variation between the response of animals in any particular litter is less than those between animals from different litters, the error of the test would be reduced. On the contrary, since this procedure would *confound* differences between doses with differences between litters, so that the effects of these two factors could not be separated in the analysis of variance, the misuse of litter mates in this way would increase the error.

If we wish to use 100 animals in a test and have not 100 animals which can reasonably be regarded as coming from a homogeneous stock, we may be justified in sorting them into groups, each group being regarded as more homogeneous than the stock as a whole. If this is done and several groups can be selected with such numbers of members that a balanced allotment to the various doses can be made in the same way as we might allot litter mates, then exactly the same procedure can be followed, for any particular animal is again as likely to be allotted to any given dose as to another. This is a better procedure than attempts to balance dosage groups by the prior selection of animals so that such a factor as the mean weight of animals in all groups shall be approximately constant. This procedure, which is not random selection, may bias the results of the test and will impair the validity of the statistical methods

which will be applied. The methods which are described in this monograph are based on the assumption that randomisation of test objects has been followed, and the absence of strict randomisation may seriously affect the reliability of any conclusions we may draw.

6.3. The Latin square

A design which has proved particularly useful as a basis for the allotment of treatments and test objects with a scheme of restricted randomisation is the Latin square. Latin squares were first used in agricultural experimentation, where field trials are conducted on plots of land which may exhibit gradients in soil fertility. A geometrical scheme for allotting treatments so that variations in soil fertility could be eliminated from the estimation of errors was necessary. A plot of land to be used in experimental work, preferably approximately square in shape, is divided into a number of similar rectangles or squares, as in a chess board, with an equal number of rows and columns (see Figure 6.1). The number of treatments to be applied must be equal to or a factor of the number of rows and columns. In the simplest case, where the number of rows, columns and treatments are the same, each treatment must be applied to a small square which falls in each different row and in each different column. This will be seen to be the case in Figure 6.1, where each capital letter represents a different treatment.

C	B	D	A	E
E	C	A	B	D
A	E	C	D	B
D	A	B	E	C
B	D	E	C	A

FIG. 6.1. A 5×5 Latin square.

Although subject to this restriction, it is still equally likely that any given treatment shall fall into any particular position in the square and thus the principles of randomisation still hold. The restrictions in design ensure, however, that since every row and every column shall contain an example of each kind of treatment, the differences between rows and columns may be eliminated in the analysis of variance and thus any systematic trend in the natural fertility of the soil from one end or side of the field to the other will be discounted.

The square used for the purposes of illustration is a 5×5 square, i.e. there are five rows and five columns. Row 1 contains examples of all treatments from A to E, and so does each other row. Thus

the natural fertility of row 1 may be compared with that of the other rows by totalling crop yields for each row. Crop yields in columns may be compared by exactly the same method. In comparing treatments the five examples of each treatment distributed over the square are similarly totalled. Hence the analysis of variance takes the following form:

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between rows	4	R	$\frac{R}{4}$
Columns	4	C	$\frac{C}{4}$
Between treatments	4	D	$\frac{D}{4}$
Random sampling (error)*	12	$T-(R+C+D)$	$\frac{T-(R+C+D)}{12}$
Total	24	T	—

The effect of treatments is estimated by comparison with the error variation based on 12 degrees of freedom after the elimination of systematic trends of fertility along the two rectangular axes of the field. Latin squares have been used in practice when comparing potencies of drugs in this physical sense; they may, for instance, be applied to the positions of test objects on a rack, such as the cages in which animals are placed during the tests; but they are frequently used in a more symbolic manner. Thus the place of rows and columns applied in the physical sense to a plot of ground may be taken by litters and order of injection, so that members of any one litter shall between them receive one of each kind of treatment, while the injections are so ordered that each sub-interval of the injection period shall also contain an example of each treatment. Any effect which litters and the time of injection may have upon response may then be eliminated in the analysis of variance by exactly the same procedure as outlined above.

Procedures for the selection of a Latin square suitable to the test in view are laid down in such books of statistical tables as that of Fisher and Yates. There are 12 possible Latin squares which have three rows and three columns, i.e. 3×3 squares, 576 possible 4×4 squares and 161,280 possible 5×5 squares, whilst the possible

* In the Latin square, it is perhaps a misnomer to call the error term random sampling, as it in fact consists of unisolated interactions, see p. 116.

number of squares greater than 5×5 increase enormously as the index of the square increases. It will be seen that there is no danger of an experimenter using up the possible combinations of such squares in the course of his work.

The restriction as to the number of doses, rows and columns which this type of design imposes upon experiments will generally be no handicap, but instead, if intelligently used, will be a great help to the research worker. There is often no cogent reason why the doses and their spacing and the number of doses or the number of animals used in biological assays are such as they are, and in planning an assay so that it shall be the most statistically convenient, all the objects of the assay can usually be attained through the utilisation of one of the restricted designs available.

We shall now consider the detail of calculation in two examples of designs employing restricted randomisation. In the first of these examples the test is so designed that in the analysis of variance we are able to segregate two possible sources of variation, other than that due to linear regression and deviation from regression, namely, differences between litters of rats and variation due to random sampling. In the second design, which is a Latin square, we are able to segregate three sources of variation not concerned with differences between doses.

6.4. A test using restrictions in design

Table 6.1 gives the weights of the uteri in mgm. of immature female rats which were injected with international standard oestrone (adapted from Bülbring and Burn, *J. Physiol.*, **85**, 320, 1935). These uterine weights are actually expressed in mgm. per 100 gm. of rat, a procedure not particularly to be recommended, as will be clear from later chapters. The figures have been taken from records of a test in which this method of measuring the response was adopted. When oestrone is injected to the immature female rat, it causes an increase in uterine weight, and this increase may be used as the response for the purpose of biological assay. The uteri of uninjected rats averaged about 30 mgm. per 100 gm. and that is all we need to know in order to decide that each level of treatment has elicited a response from the group receiving it. Members of eight litters of rats were so distributed that each dosage group of eight rats contained a member from each litter. We will suppose that litter mates were allotted to doses strictly at random and that we may validly apply the analysis of variance.

Since each litter of three animals was injected with each of the three different doses, the total uterine weight for each litter is a measure of differences between litters, and this is listed under the heading T_l in the Table. Since each dosage group includes a

TABLE 6.1

THE RESPONSE OF RAT UTERI (WEIGHT IN MGM.) TO INJECTED OESTRONE IN EIGHT LITTERS OF THREE RATS, ONE MEMBER ON EACH DOSE

Litter	Dose of oestrone			
	0·2 μg.	0·4 μg.	0·8 μg.	Totals, T_l
1	106	116	145	367
2	72	88	135	295
3	42	68	115	225
4	64	111	136	311
5	70	111	133	314
6	56	68	85	209
7	42	63	87	192
8	65	70	150	285
Totals, T_p	517	695	986	2,198 (T)

ST_p^2=1,722,510; n_p=8
ST_l^2= 629,426; n_l=3
T^2=4,831,204; n=24

SY^2=226,218·0	$\frac{ST_p^2}{n_p}$=215,313·8	$\frac{ST_l^2}{n_l}$=209,808·7
$\frac{T^2}{n}$=201,300·2	$\frac{T^2}{n}$=201,300·2	$\frac{T^2}{n}$=201,300·2
Difference, Sy^2= 24,917·8	$Sn_p\bar{y}_p^2$= 14,013·6	$Sn_l\bar{y}_l^2$= 8,508·5

member from each litter, differences between doses are similarly measured by the corresponding dosage totals, T_p. The correction for the mean response, $\bar{Y}$, throughout the calculations will be the sum of all responses, T, multiplied by the mean response. This is equivalent to $\frac{T^2}{n}$, where n is the total number of animals. Sums of squares are shown at the bottom of Table 6.1: Sy^2 is the total sum of squares, $Sn_p\bar{y}_p^2$ is the sum of squares for the means of dosage groups and $Sn_l\bar{y}_l^2$ is the sum of squares for the means of litters, where n_p=8 and n_l=3, the numbers of animals in the dosage groups and litters respectively.

We then calculate the regression coefficient in Table 6.2. The doses were 0·2, 0·4 and 0·8 μg. and $\bar{X}$, the logarithm of the $\frac{\text{dose}}{2}$ to base 2, takes the convenient numbers −1, 0, +1, used as coefficients for the log dose. Against these we list $\bar{Y}_p$, the mean response of each group $= \frac{T_p}{8}$, and calculate the sums of squares and products as in previous analyses. Since $\bar{X}=0$, the only correction for the

TABLE 6.2

CALCULATION OF REGRESSION FOR THE DATA OF TABLE 6.1

Dose in 0·2 μg.	Coefficient for log dose, $\bar{X}_p$	Response, $\bar{Y}_p$	$\bar{X}_p^2$	$\bar{X}_p\bar{Y}_p$	$\bar{Y}_p^2$
1	−1	64·625	1	−64·625	4,176·4
2	0	86·875	0	0·000	7,547·3
4	1	123·250	1	123·250	15,190·6
Totals	0	274·750	2	58·625	26,914·3
Means	0	91·583	—	—	—
Correction for mean			0	0·0	25,162·4
Sums of squares and products			2	58·625	1,751·9

$$b = \frac{S\bar{x}_p\bar{y}_p}{S\bar{x}_p^2} = \frac{58{\cdot}625}{2} = 29{\cdot}313$$

mean which differs from zero is that applying to $\bar{Y}_p^2$, which is not in fact needed for the calculation of regression, but which we have included for the purposes of checking the calculations. We then proceed to isolate the sum of squares attributable to linear regression exactly as in previous examples, and by subtracting this from $Sn_p\bar{y}_p^2$ we obtain the sum of squares attributable to deviations from regression, as in Table 6.3. There is only one degree of freedom associated with each of these first two sums of squares, since there are only three dosage groups. The sum of squares attributable to variation between litters is $Sn_l\bar{y}_l^2$, with which are associated seven degrees of freedom, one less than the number of litters, and the residual sum of squares is determined by subtracting these three sums of squares from Sy^2, and with it are associated the remaining 14 degrees of freedom, giving us a total of 23 degrees of freedom. This residual sum of squares, which will be used as our

term for error, is that attributable to random sampling within litters.

The sum of squares between doses is 14,013·6, which, divided by 8, should check with $S\bar{Y}_p^2$ in Table 6.2. This latter sum is 1.751·9 and $\frac{14{,}013{\cdot}6}{8}=1{,}751{\cdot}7$, which agrees within the limits of error of the calculations. The sum of squares attributable to linear regression, which is $\frac{(\bar{X}_p\bar{Y}_p)^2 n_p}{S\bar{X}_p^2}$, since n_p is constant (this equals $\frac{n_p(S\bar{x}_p\bar{y}_p)^2}{S\bar{x}_p^2}$ since $S\bar{X}_p=0$), is 13,747·6, leaving us 266·0 as the sum of squares for the deviations from regression. The mean square for random sampling, 171·7, is to be compared with the three sums of squares above it in Table 6.3, whence we get the series of values of F listed

TABLE 6.3

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 6.1

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square	F	P
Between doses	$Sn_p\bar{y}_p^2$					
A. Linear regression	$\frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p^2}$	1	13,747·6	13,747·6	80·3	<0·001
B. Deviations from regression	$Sn_p\bar{y}_p^2-A$	1	266·0	266·0	1·6	>0·05
C. Between litters	$Sn_t\bar{y}_t^2$	7	8,508·5	1,215·5	7·1	<0·01
Random sampling within litters (=error)	$Sy^2-(A+B+C)$	14	2,395·7	171·7	—	—
Total	Sy^2	23	24,917·8	—	—	—

in the table and the corresponding probabilities, P, associated with them. We see that, while the significance of the regression is beyond doubt, there are no significant deviations from this regression. The variation between litters is also significantly greater than that within them and we have thus eliminated in the design of the test an important source of variation. Had we not so designed the test as to rule out the inter-litter variation, our sum of squares for error would include that between litters and would thus be 8,508·5+2,395·7 (see Table 6.3)=10,904·2, with which would

be associated 21 degrees of freedom, giving us a mean square of 519·2, which is three times as large as the mean square for random sampling within litters, and thus our test would have been conducted with only one-third of the precision obtainable by the more adequate statistical design. To obtain the regression equation, we substitute in the formula $E-\bar{Y}=b(X-\bar{X})$, whence we get

$$(E-91{\cdot}583)=29{\cdot}313(X-0), \text{ or } E=91{\cdot}583+29{\cdot}313X,$$

where X is $\log_2\left(\frac{\text{dose}}{2}\right)$.

6.5. A test using the Latin square

Table 6.4 gives the protocols of a test using a 5×5 Latin square. It is adapted from material relating to the assay of a principle of the

TABLE 6.4

RELATIONSHIP OF DOSE OF THYROTROPHIN TO THE WEIGHT OF THE THYROID GLANDS OF GUINEA-PIGS, USING A LATIN SQUARE DESIGN

Strain	Cage 1	2	3	4	5	Totals, T_s
1	C 65	E 85	A 57	B 49	D 79	335
2	E 82	B 63	D 77	C 70	A 46	388
3	A 73	D 68	C 51	E 76	B 52	320
4	D 92	C 67	B 63	A 41	E 68	331
5	B 81	A 56	E 99	D 75	C 66	377
Totals, T_c	393	339	347	311	311	1,701 (=T)

Dose	A	B	C	D	E	
Totals, T_p	273	308	319	391	410	1,701 (=T)
Means, $\bar{Y}_p$	54·6	61·6	63·8	78·2	82·0	

	$SY^2=120{,}719{\cdot}00$	$\frac{ST_s^2}{n_s}=116{,}111{\cdot}80$
	$\frac{T^2}{n}=115{,}736{\cdot}04$	$\frac{T^2}{n}=115{,}736{\cdot}04$
Difference,	$Sy^2=\quad 4{,}982{\cdot}96$	$Sn_s\bar{y}_s^2=\quad 375{\cdot}76$
	$\frac{ST_c^2}{n_c}=116{,}644{\cdot}20$	$\frac{ST_p^2}{n_p}=118{,}427{\cdot}00$
	$\frac{T^2}{n}=115{,}736{\cdot}04$	$\frac{T^2}{n}=115{,}736{\cdot}04$
Difference,	$Sn_c\bar{y}_c^2=\quad 908{\cdot}16$	$Sn_p\bar{y}_p^2=\quad 2{,}690{\cdot}96$

pituitary gland, thyrotrophin, which causes growth of the thyroid gland in mammals (Rowlands and Parkes, *Biochem. J.*, **28**, 1829, 1934). Guinea-pigs of about 200 gm. in weight are injected for several days with an extract of cattle pituitary glands and are then killed and the weights of their thyroid glands determined. The thyroid glands of uninjected guinea-pigs of this weight averaged 31 mgm. and hence we may be satisfied that the lowest of the doses given has produced a response, since the mean weight of the thyroids in the group receiving it was 54.6 mgm. We will suppose that five cages, each containing five pigs, were placed along a rack in the animal house and that each cage contained one animal from each of five different strains. (It would be difficult to obtain five sets of five litter-mates, since guinea-pigs rarely produce so many at a birth.)

Having selected a 5×5 Latin square at random, we allot each of the five doses to positions A—E in the square, so that each cage will contain one animal receiving each dose, as well as one animal from each strain, and one animal from each strain will also receive each of the five doses. Thus each dosage group will contain one member from each strain and at the same time one member from each cage. This design, as explained above, enables us to separate the variation attributable to differences between cages, differences between strains, differences between doses (which will be further divided into that due to linear regression and to departures from regression) and to random sampling. We work in totals, as in Table 6.1, and list the four sums of squares at the bottom of Table 6.4. Sy^2 is the total sum of squares, $Sn_c\bar{y}_c^2$ is the sum of squares between cages and $Sn_p\bar{y}_p^2$ the sum of squares between doses. The corresponding totals will be indicated with the same subscript as that of the group to which they belong.

In Table 6.5 we calculate the regression of the totals of groups on dose. This is a more rapid method and we use it now that the methods of dealing with regression should be well understood. The doses in the groups A—E are shown in Table 6.5. They were so arranged that each dose is 1·5 times the previous one. We determine $\bar{X}_p$ by taking the logarithms of (dose/4·5) to the base 1·5, which gives us a symmetrical arrangement of coefficients for log doses so that $S\bar{X}_p = 0$ and there will be no correction for the mean. We do not include a check column in the table, but this can be added by the reader if he desires to check deviations from regression, etc. The calculation of the regression line by this

scheme is thus extremely simple, but it should be remembered that we want to finish up with a regression relating the mean response to dose. To obtain this we must divide the regression coefficient for

TABLE 6.5

REGRESSION OF THYROID GLAND WEIGHT ON DOSE OF THYROTROPHIN

	Dose in mgm.	Coefficient for log dose $\bar{X}_p=\log_{1\cdot5}$ (dose/4·5)	T_p	$\bar{X}_p^2$	$\bar{X}_pT_p$
A	2·00	−2	273	4	−546
B	3·00	−1	308	1	−308
C	4·50	0	319	0	0
D	6·75	1	391	1	391
E	10·13	2	410	4	820
Totals		0	1,701	10	357

For means $b=\dfrac{357}{10\times5}=7\cdot14$ $\qquad \bar{Y}=\dfrac{T}{25}=68\cdot04$

$$E-68\cdot04=7\cdot14(X-0)$$
$$E=68\cdot04+7\cdot14X$$

totals by the number of animals in each group. The calculation is shown at the bottom of the Table and we obtain the regression line $E=68\cdot04+7\cdot14X$, in which X is $\log_{1\cdot5}\left(\dfrac{\text{dose}}{4\cdot5}\right)$.

We may wish to use this regression in conjunction with common logarithms, and it may be transformed so that $X'=\log_{10}(\text{dose})$ by means of the identity:

$$\log_N\left(\frac{A}{B}\right)=\frac{\log_{10}A-\log_{10}B}{\log_{10}N}$$

whence, in this case,

$$\begin{aligned}E&=68\cdot04+7\cdot14\frac{X'-\log_{10}4\cdot5}{\log_{10}1\cdot5}\\&=68\cdot04+40\cdot547X'-26\cdot49\\&=41\cdot55+40\cdot55X'.\end{aligned}$$

The regression is illustrated in Figure 6.2.

In the next chapter we shall be introduced to a general method of computation for dose-response lines which eliminates even this amount of work, as long as the doses have been properly chosen,

and is applicable to cases where a transformation such as that shown above would be rather more complicated.

The sum of squares attributable to linear regression may be calculated directly from the last two columns of Table 6.5. It is $\frac{(S\bar{X}_p T_p)^2}{n_p S\bar{X}_p{}^2}$. This formula is the same as the one given in the analysis of variance in Table 6.6. From the analysis of variance

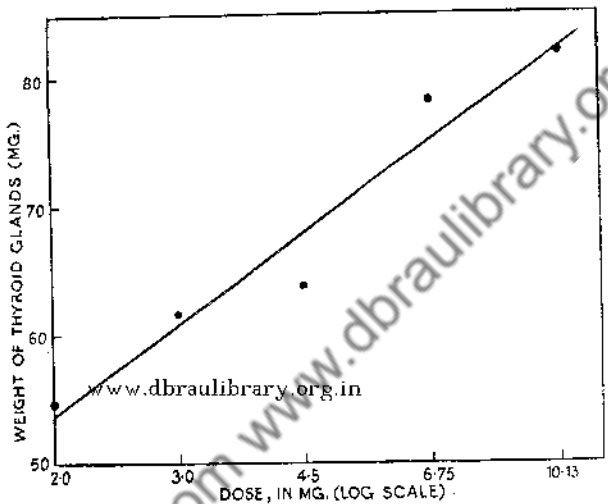


FIG. 6.2. The response of the thyroid glands of rats to thyrotrophin (Table 6.4).

given in this Table it will be seen that, as it so happens, the differences between strains and cages are not significant, nor are the deviations from regression, while the significance of the regression coefficient is without doubt. Nevertheless, the elimination of the variation attributable to two sources (strains and cages) the mean squares for which do not differ significantly from the mean square due to random sampling has reduced the estimate of the mean square for error from 114·6 to 84·0. In many other instances the reduction would be far greater, as was the reduction due to the elimination of inter-litter differences in the last example.

In connection with the reduction that may occur in the mean square for error by the elimination of various factors, it should be noted that an excessive reduction in the number of degrees of freedom available for the estimation of error is to be avoided, unless accompanied by the elimination of a significant amount of

variation. The reduction seen in the last example, from an error mean square of 114·6 to 84·0, was accompanied by a reduction in degrees of freedom from 20 to 12. The total effect is to reduce the estimated error, although not significantly, since with 20 degrees of freedom, t at, for example, $P=0{\cdot}05$, is 2·086, and ts is 22·3

TABLE 6.6

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 6.4

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square	F	P
A. Between strains	$Sn_s\bar{y}_s^2$	4	375·76	93·9	1·1	>0·05
B. Between cages	$Sn_c\bar{y}_c^2$	4	908·16	227·0	2·7	>0·05
Between doses	$Sn_p\bar{y}_p^2$					
C. Linear regression	$\frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p^2}$	1	2,548·98	2,549·0	30·3	<0·001
D. Deviations from regression	$Sn_p\bar{y}_p^2-C$	3	141·98	47·3	0·56	>0·05
E. Random sampling (=error)	$Sy^2-(A+B+C+D)$	12	1,008·08	84·0	—	—
Total	Sy^2	24	4,982·96	—	—	—
$A+B+E$	$Sy^2-Sn_p\bar{y}_p^2$	20	2,292·00	114·6	—	—

approx.; with 12 degrees of freedom t at $P=0{\cdot}05$ is 2·179 and ts is 20·0. If the same mean squares had occurred in the course of reducing the degrees of freedom available for the estimation of error from, say, 12 to 4, the estimated error of the test would thereby be increased from 23·3 to 25·4 ($P=0{\cdot}05$).

Thus, if preliminary tests indicate that a particular source of variation (or supposed source thereof) is in fact unimportant, there may be nothing to be gained in future by continuing to segregate it, with the accompanying loss of degrees of freedom. If the design of the test has, however, been such as to segregate a particular source of variation, it is not usually wise to recombine an apparently non-significant sum of squares with the error sum of squares merely in the light of the observed result, as this procedure will bias the estimate of error in the long run. If we design a test so that a particular set of degrees of freedom is allotted for error, only this set should be used whether or not other sources of variation prove to be significant in the particular test analysed.

CHAPTER 7

POLYNOMIAL COEFFICIENTS

7.1. Introducing the coefficients

The calculation of dose-response relationships is so much easier if the doses are equally spaced on a logarithmic scale and the numbers of observations per dose are equal, that we shall consider only this case for the present. There is, after all, rarely any point in failing to space doses and to allot observations in such a way as to save trouble and to gain the maximum of information from results. It is not always possible to do this, but is more often possible than a survey of the literature would suggest.

When the log doses are equally spaced, we make further use of the coefficients for log dose we have already met, which are small whole numbers bearing the same relation to each other as that between the differences of each log dose from the mean log dose. These are *polynomial coefficients*. By means of these coefficients it is easy to examine the possible curvature of a log dose-response line and to determine how well the supposed relationship between log dose and response fits the data. The exact nature of discrepancies may also be examined, but we shall not concern ourselves with more than an isolation of the sums of squares attributable to linear or quadratic regression—i.e. those accounted for by fitting to the data a straight or a singly curved line of the form $Y=a+bX$ or $Y=a+bX+cX^2$ respectively.

Two points can always be fitted by a straight line and three points by a singly curved line. As more points are considered, we may fit them by curves of higher degrees. In general, n points can be completely linked by a curve of the $(n-1)$th degree (involving terms up to x^{n-1}). Each successive term in the equation

$$Y=a+bX+cX^2+dX^3+ \ . \ . \ . \ nX^{n-1}$$

can be examined by the use of polynomial coefficients to see whether it significantly improves the fit of a curve to a series of points.

The coefficients applicable to an examination of the linear and quadratic terms are shown in Table 7.1, for from three to eight doses. These coefficients are *orthogonal*, which means that with each single

degree of freedom a set of corresponding coefficients may be chosen so that independent estimates are made of the importance of each source of variation. The scheme for computation, also outlined in Table 7.1, is as follows. The sum of the n_p responses on each dose is entered at the foot of the relevant column; this is T_p. The coefficients are represented by the letter k, and the divisor, n_pSk^2, is the number of responses per group multiplied by the sum of the squares of coefficients in each row. The sum of products for each row is ascertained by multiplying each T_p by the corresponding coefficient—with due regard for sign—and adding the results. Tabulation is conveniently made as in Table 7.1, with a final column for the mean square, or variance, which is the square of a sum of products divided by the corresponding divisor, or

$$\frac{(SkT_p)^2}{n_pSk^2}$$

TABLE 7.1

POLYNOMIAL COEFFICIENTS AND SCHEME FOR ISOLATING THE SUMS OF SQUARES ATTRIBUTABLE TO THE LINEAR AND QUADRATIC TERMS OF A REGRESSION LINE

Term	Coefficients (k) for doses								Divisor	Sum of products	Mean square
	1	2	3	4	5	6	7	8	n_pSk^2	SkT_p	$\frac{(SkT_p)^2}{n_pSk^2}$
Linear	−1	0	+1	—	—	—	—	—	$2n_p$		
Quadratic	+1	−2	+1	—	—	—	—	—	$6n_p$		
Linear	−3	−1	+1	+3	—	—	—	—	$20n_p$		
Quadratic	+1	−1	−1	+1	—	—	—	—	$4n_p$		
Linear	−2	−1	0	+1	+2	—	—	—	$10n_p$		
Quadratic	+2	−1	−2	−1	+2	—	—	—	$14n_p$		
Linear	−5	−3	−1	+1	+3	+5	—	—	$70n_p$		
Quadratic	+5	−1	−4	−4	−1	+5	—	—	$84n_p$		
Linear	−3	−2	−1	0	+1	+2	+3	—	$28n_p$		
Quadratic	+5	0	−3	−4	−3	0	+5	—	$84n_p$		
Linear	−7	−5	−3	−1	+1	+3	+5	+7	$168n_p$		
Quadratic	+7	+1	−3	−5	−5	−3	+1	+7	$168n_p$		
Total of responses on each dose $=T_p$											

When three dosage groups are employed, the sums of squares for the linear and quadratic terms, with each of which is associated one degree of freedom, should add up to the sum of squares for differences between doses, since only two degrees of freedom are involved. This is equivalent to saying that the responses to three doses can be joined by a parabola of the form

$$Y=a+bX+cX^2$$

However, the curve is only of interest if the mean square for the quadratic term exceeds that for random sampling significantly —otherwise a straight line describes the relationship as well as does a curved line. With more than three doses there will be a residual sum of squares associated with the remaining degrees of freedom, which can be determined by subtraction of the sums of squares corresponding to the linear and quadratic terms from the sum of squares for all dosage effects.

7.2. Use of the coefficients

We may, as an exercise, calculate the linear and quadratic terms for the two sets of data in Chapter 6. Table 7.2 shows the re-

TABLE 7.2

ANALYSIS OF THE DOSE-RESPONSE RELATION OF TABLE 6.2 BY POLYNOMIAL COEFFICIENTS

Term	Coefficients for log dose (k)			Divisor	Sum of products	Mean square $\frac{(SkT_p)^2}{n_pSk^2}$
	0·2 μg.	0·4 μg.	0·8 μg.	n_pSk^2	SkT_p	
Linear	−1	0	+1	16	469	13,747·6
Quadratic	+1	−2	+1	48	113	266·0
Total of responses, T_p	517	695	986			

calculation by use of polynomial coefficients of the sources of variation A and B in Table 6.3.

Table 7.3 shows the calculation of the mean squares associated with the linear and quadratic terms of the regression

TABLE 7.3

ANALYSIS OF THE DOSE-RESPONSE RELATION OF TABLE 6.5 BY POLYNOMIAL COEFFICIENTS

Term	Coefficients for log dose (k)					Divisor	Sum of products	Mean square $\frac{(SkT_p)^2}{n_pSk^2}$
	2·00	3·00	4·50	6·75	10·13	n_pSk^2	SkT_p	
Linear	−2	−1	0	+1	+2	50	357	2,549·0
Quadratic	+2	−1	−2	−1	+2	70	29	12·0
Residual	—	—	—	—	—	—	—	65·0
Totals of responses, T_p	273	308	319	391	410			

in Table 6.6. That for the quadratic term is identical with the mean square for deviations from regression. The linear term is as before, and the insignificant contribution of the quadratic

term—predictable from the previous finding that the mean square associated with deviations from regression was smaller than that for error—is also determined. The residual mean square, with which are associated two degrees of freedom, is obtained by subtraction from the previously-determined sum of squares attributable to dosage effects. It could be obtained by calculating the sum of squares for cubic and quartic terms, which together account for all the possible departures from linearity not included in the quadratic term.

We used coefficients based on logarithms to base 2 and base 1·5 in the analyses of Chapter 6, because of their mathematical convenience. We can abandon such devices now that we have adopted polynomial coefficients, which are identical for the linear term with our log doses. This is not an accident, but a feature of the convenience at which we aimed. What should we have done if the number of doses had been even? Then we should have had to interpolate imaginary doses between the real ones and get a symmetrical distribution of log doses by the following steps:

Actual doses	1		2		4		8
Imaginary doses		$\sqrt{2}$		$2\sqrt{2}$		$4\sqrt{2}$	
$\text{Log}_{\sqrt{2}}$ dose	0	(1)	2	(3)	4	(5)	6
$\text{Log}_{\sqrt{2}}\left(\frac{\text{dose}}{2\sqrt{2}}\right)$	−3	(−2)	−1	(0)	1	(2)	3

Whence we arrive at the polynomial coefficients for the linear term for four dosage groups: −3, −1, +1, +3.

7.3. Calculation of the regression line using polynomial coefficients

The simple rule for determining the slope, b, of a dose-response line when polynomial coefficients are used is:

$$b=\frac{SkT_p}{In_pSk^2} \text{ for an odd number of doses}$$

$$\text{and } b=\frac{2SkT_p}{In_pSk^2} \text{ for an even number of doses}$$

where I=the interval in logarithms (or any other unit) between successive doses.

The position of the line is determined from the equation

$$(E-\bar{Y})=b(X-\bar{X})$$

where $\bar{Y}$ and $\bar{X}$ are the mean response and mean log dose to any base.

After conversion to base 10, we obtained the regression equation

$$E=41{\cdot}55+40{\cdot}55X$$

for the data of Tables 6.4–6.6.

From Table 7.3,

$$b=\frac{357}{0{\cdot}1761\times 50}=40{\cdot}55$$

whence

$$E-68{\cdot}04=40{\cdot}55(X-0{\cdot}6532)$$
$$E=41{\cdot}55+40{\cdot}55X.$$

CHAPTER 8

COVARIANCE

8.1. Correcting for initial or concomitant differences between test objects

In calculating the dose-response line for the responses of rats injected with thyrotrophin in Chapter 6 we were able in the design of the test to allow for the elimination of various possible sources of error. We did not allow for the correction of any concomitant variation exhibited by the animals themselves. The weight of a guinea-pig may determine in part the weight of its thyroid gland, whether this gland has increased in response to stimulation during the test or not. Guinea-pigs in such a test are normally chosen so as to vary as little as possible in weight, but it is rarely feasible to choose large groups of animals of so nearly the same weight that differences between them in that respect can be neglected. We shall now investigate a method by which the influence of the weight of the animal, or any other variable in which we may be interested, may be examined. This is the method of *covariance analysis*.

Table 8.1 shows a series of coded body weights which have been assigned to the pigs in the corresponding positions of Table 6.4. If the weights of the original pigs had varied between 150 and 250 gm., then for the examination of the effect of weight on response it would be sufficient to take the weight of each animal to the nearest 5 gm. and to subtract 150 gm. from each weight and divide by 5. This would give us such a series of numbers for easy working as in Table 8.1.

Covariance is computed by an expansion of the methods used in the analysis of variance. For each sum of squares or products derived from X and Y, we must compute corresponding sums of squares of W, the weights of the animals, or, as in this case, a coded figure representing them. A third set of values, the sums of the products WY, must also be computed. The rule for forming this third set is that at each and every stage where a number is squared in calculating sums of squares for W and Y, the paired values of W and Y are multiplied together to obtain the sum of products, WY. Thus we run the calculations in Table 8.1 in parallel with

those of Table 6.4, computing the various totals indicated by t (not to be confused with "Student's" "t") for the various restrictions in design and calculating the sums of squares and sums of products which are shown at the bottom of the Table. From SW^2, $\frac{St_c^2}{n_c}$, etc., we subtract $\frac{t^2}{n}$ to obtain Sw^2, Sw_c^2, etc. From SWY, $\frac{ST_ct_c}{n_c}$, etc., we subtract $\frac{Tt}{n}$ to obtain Swy, $S\bar{w}_c\bar{y}_c$, etc. From these we then build up an analysis of covariance similar in design to the analysis of variance in Table 6.6.

TABLE 8.1

DATA FOR THE ADJUSTMENT OF THE MATERIAL IN TABLE 6.4 FOR VARIATION IN THE WEIGHT OF GUINEA-PIGS

Strain	Cage					Totals, t_s
	1	2	3	4	5	
1	C 10	E 14	A 9	B 12	D 11	56
2	E 0	B 18	D 15	C 20	A 10	63
3	A 19	D 2	C 6	E 9	B 1	37
4	D 18	C 5	B 8	A 3	E 6	35
5	B 11	A 6	E 18	D 9	C 12	56
Totals, t_c	56	43	55	53	40	247 ($=t$)
Dose Totals, t_p	A 47	B 49	C 51	D 53	E 47	

$SW^2 = 3{,}239{\cdot}00$
$St_c^2/n_c = 2{,}483{\cdot}80$
$St_s^2/n_s = 2{,}567{\cdot}00$
$St_p^2/n_p = 2{,}445{\cdot}80$
$t^2/n = 2{,}440{\cdot}36$

$Sw^2 = 798{\cdot}64$
$Sn_cw^2 = 43{\cdot}44$
$Sn_sw^2 = 126{\cdot}64$
$Sn_pw_p^2 = 5{\cdot}44$

$SWY = 17{,}645{\cdot}00$
$ST_ct_c/n_c = 16{,}918{\cdot}60$
$ST_st_s/n_s = 16{,}918{\cdot}20$
$ST_pt_p/n_p = 16{,}837{\cdot}00$
$Tt/n = 16{,}805{\cdot}88$

$Swy = 839{\cdot}12$
$Sn_c\bar{w}_c\bar{y}_c = 112{\cdot}72$
$Sn_s\bar{w}_s\bar{y}_s = 112{\cdot}32$
$Sn_p\bar{w}_p\bar{y}_p = 31{\cdot}12$

We must also determine the slope of the regression of thyroid weight on body weight. For this we use Sw^2 and Swy from the error row only, as entries in this row are free from variation attributable to differences between doses and restrictions in design. The slope, $b_w = \frac{Swy}{Sw^2}$, is then used to correct each Sy^2 for variations in W.

8.2. The analysis of variance and covariance

Table 8.2 shows the new analysis of variance, omitting the examination of regression, which will be made subsequently, using polynomial coefficients. The slope, b_w, is $\frac{582{\cdot}96}{623{\cdot}12}$, or 0·93555, and is used to correct each Sy^2 to form an "adjusted" Sy^2, denoted by Sy_a^2, where

$$Sy_a^2 = b_w^2 Sw^2 - 2b_w Swy + Sy^2$$
$$= 0{\cdot}87525 Sw^2 - 1{\cdot}8711 Swy + Sy^2$$

These coefficients are shown for convenience at the bottom of Table 8.2. They are those of the expansion $(y - b_w w)^2 = y^2 - 2b_w wy + b_w^2 w^2$, since we are comparing quantities $(y - b_w w)$. Since the statistic, b_w, has been computed from the row for random sampling, we have one less degree of freedom in this row, namely, 11 instead of 12. With this exception, the mean squares are determined as before, and F calculated from them. The elimination of the one degree of freedom associated with covariance represents a variance of 1,008·08 − 462·69 = 545·39. This is a highly significant variance compared with the variance for error, and in fact has reduced the variance associated with random sampling to almost exactly one-half of its former value, and its elimination has increased the precision of the experiment twofold.

TABLE 8.2

ANALYSIS OF COVARIANCE FOR THE DATA OF TABLE 6.4

Source of variation	Degrees of freedom	Sums of squares and products			Adjusted y^2		F	P
		Sw^2	Swy	Sy^2	Sum of squares	Mean square		
Between strains	4	126·64	112·72	375·76	275·69	68·9	1·64	>0·05
Between cages	4	43·44	112·32	908·16	736·02	184·0	4·37	<0·05
Between doses	4	5·44	31·12	2,690·96	2,637·49	659·4	15·70	<0·001
Random sampling	11	623·12	582·96	1,008·08	462·69	42·1	—	—
Total	23	798·64	839·12	4,982,96	—	—	—	—
Coefficients for adjusting Sy^2	—	0·87525	−1·8711	1·0				

8.3. Computation of regression with covariance

The same methods are applied in computing the linear regression of response on dose, and deviations from it. Sums of body

weights, t_p, are added to the polynomial analysis, and two columns of products are formed, Skt_p and SkT_p (Table 8.3).

TABLE 8.3

COVARIANCE ANALYSIS OF THE DOSE-RESPONSE LINE FOR THE DATA OF TABLE 8.1

Term	Coefficients for log dose					Divisor	Sums of products		Variances and covariances			Adjusted
	2·00	3·00	4·50	6·75	10·13	n_pSk^2	Skt_p	SkT_p	Sw^2	Swy	Sy^2	Sy_a^2
Linear	−2	−1	0	+1	+2	50	4	357	0·32	28·56	2,549·0	2,495·8
Quadratic	+2	−1	−2	−1	+2	70	−16	29	3·66	−6·63	12·0	27·6
Residual	—	—	—	—	—	—	—	—	—	—	—	57·0
						Coefficients for adjusting Sy^2			0·87525	−1·8711	1.	
Totals: t_p	47	49	51	53	47							
T_p	273	308	319	391	410							
Mean square for random sampling (error)												42·1

Three columns of variances and covariances, Sw^2, Swy and Sy^2, are required, such that:

$$Sw^2=\frac{(Skt_p)^2}{n_pSk^2} \qquad Swy=\frac{(Skt_p)(SkT_p)}{n_pSk^2} \qquad Sy^2=\frac{(SkT_p)^2}{n_pSk^2}$$

Correction for variations in W is made by multiplying Sw^2 by $b_w{}^2$ and Swy by $-2b_w$ and proceeding as before. The adjusted Sy^2, or $Sy_a{}^2$, is shown in the last column of Table 8.3. From this Table, it is again apparent that a linear regression adequately describes the relationship between dose and response.

The adjusted value, $Sy_a{}^2$, for the linear term determines the new slope of the dose-response line. It is designated by the symbol B^2 and related to the slope, b, such that:

$$b=\frac{B}{I\sqrt{n_pSk^2}} \text{ for an odd number of doses, and}$$

$$b=\frac{2B}{I\sqrt{n_pSk^2}} \text{ for an even number of doses,}$$

where, it will be recalled, I is the log dose interval, and B has the same sign as SkT (Bliss and Marks, *Quart. J. Pharmacol.*, **12**, 82, 1939).

In Table 8.3, $B^2=2{,}495{\cdot}8$, whence

$$b^2=\frac{2{,}495{\cdot}8}{I^2\times 50}=\frac{2{,}495{\cdot}8}{0{\cdot}1761^2\times 50}$$

and $b=40{\cdot}11$—very nearly the same as before.

The corrected equation is then:

$$E-68{\cdot}04=40{\cdot}11(X-0{\cdot}6532)$$
$$E=41{\cdot}84+40{\cdot}11X.$$

8.4. Correction of mean responses

Each T_p, the total response per dosage group, may also be corrected by covariance, simply by listing the deviations in t_p, the total body weight per group, from $\frac{t}{N}$, their mean, where $N=$ number of arrays, and multiplying each deviation by b_w. This is shown in Table 8.4. The quantity $b_w\left(t_p-\frac{t}{N}\right)$ is subtracted from the total response per group, T_p, and gives the corrected total response.

TABLE 8.4

CORRECTION OF MEAN RESPONSE AT EACH DOSE FOR WEIGHTS OF ANIMALS

Dose in mg.	t_p	Deviation of t_p from t/N	T_p	Corrected total response	Corrected mean response
2·0	47	−2·4	273	275·25	55·05
3·0	49	−0·4	308	308·37	61·67
4·5	51	1·6	319	317·50	63·50
6·75	53	3·6	391	387·63	77·53
10·13	47	−2·4	410	412·25	82·45
			Total	1,701·00	

8.5. Reduced sums of squares

In the above analysis we found that after correction by covariance, the differences in response between cages appeared significant, with an F of 4·37, corresponding to a P of between 0·05 and 0·01. However, adjusted values of Sy^2 are not suitable for exact tests of significance, because the slope of the lines on which they are based is subject to error. For the precise test, a slope relevant to the particular comparison to be critically examined is required, based on the individual comparison plus random sampling. The statistic to be calculated is the *reduced* Sy^2, which is always less than the adjusted Sy^2. Therefore, if an adjusted Sy^2 indicates that significant, but not highly significant, differences exist among a set of mean responses segregated in design, it is always possible that the more critical test will show no significant differences. If the adjusted Sy^2 indicates no significant differences, the reduced Sy^2 need not be calculated, as it will always show an even smaller effect.

TABLE 8.5

CALCULATION OF REDUCED Sy^2 FOR VARIATION BETWEEN CAGES IN TABLE 8.2

Source of variation	Degrees of freedom	Sums of squares and products			Reduced Sy^2		F	P
		Sw^2	Swy	Sy^2	Sum of squares	Mean square		
Between cages	4	43·44	112·32	908·16	728·31	182·1	4·32	<0·05
Random sampling	11	623·12	582·96	1,008·08	462·69	42·1	—	—
Total	15	666·56	695·28	1,916·24	1,191·00			

To calculate a reduced Sy^2, the sums of squares and products, Sw^2, Swy and Sy^2, are added to the corresponding sums in the "random sampling" row. Calling these sums $S'w^2$, $S'wy$ and $S'y^2$, the reduced Sy^2 for any particular combination is $S'y^2 - \frac{(S'wy)^2}{S'w^2}$.

In order to extract the reduced Sy^2 for the treatment or restriction in design alone, the reduced Sy^2 for random sampling only (identical with the adjusted Sy^2) is subtracted from the reduced Sy^2 for the combination of effect plus random sampling.

We do this for the responses between cages as against error in Table 8.5, and see that the reduced Sy^2 for differences between cages is still significant.

8.6. Multiple covariance

It is possible to extend the method of covariance to deal with more than one set of concomitant observations simultaneously. Thus, the age of the guinea-pigs as well as their weight might have been entered in Table 8.1, and its effects, if any, eliminated as were those of weight. However, the computation becomes increasingly laborious and difficult to check as more variates are added, and it would frequently be easier to increase precision by enlarging the groups initially or by exercising greater control over the degree of variation shown by the experimental material. Multiple covariance would rarely be suitable in routine assays—even the valuable increase in precision that may often be obtained by simple covariance analysis may be offset by the labour involved in making additional measurements, if these are not already made as a routine, and in additional calculation. The methods of increasing precision to be adopted must be left to individual choice.

CHAPTER 9

PREDICTING FROM DOSE-RESPONSE LINES AND PLANNING ASSAYS

9.1. Errors of estimation

The variance of the estimate of the position of a log dose-response line is determined by the variance attributable to error (random sampling) in the analysis of variance. Thus, in the line

$$(Y-\bar{Y})=b(X-\bar{X}) \quad \text{or} \quad Y=a+bX$$

$$V\bar{Y}=\frac{Ve}{n}$$

and if $\bar{X}=0$, $V\bar{Y}=Va$

where n is the total number of observations and Ve is the variance associated with random sampling.

The variance of the estimate of the slope, b, is determined by the error variance and also by the spacing of the dosage groups in

$$Vb=\frac{Ve}{Sn_p\bar{x}_p^2}$$

If polynomial coefficients have been used, Vb is $\dfrac{Ve}{I^2Sn_pk_p^2}$ or $\dfrac{2Ve}{I^2Sn_pk_p^2}$ for an odd and even number of doses respectively.

When we have corrected for covariance, the mean square for error is the adjusted mean square, which will have been reduced in magnitude according to the degree to which allowance for the concomitant factor increases the precision of the test.

9.2. Planning an assay

Once a dose-response line has been established, we shall wish to plan further work in which the potencies of other preparations will be compared with that of the standard. If the line is sufficient to describe the relation between dose and response, we can predict the probable limits of error within which we shall be able to assay potencies under various conditions. If the departures from linearity are significant, we are faced with two alternatives. We can allow for the discrepancy by using the mean square for

departures from linearity instead of the mean square for error. This will not be a safe procedure unless a large number of dosage groups has been used, since the mean square will otherwise be based on too few degrees of freedom and be subject to large errors. Instead, it would be better to experiment with the data to see if a different way of expressing them gives a linear relationship, or to note that perhaps the largest dose produced an insufficient increase in response to fall on the same line as the others, and thus to limit the range of assay to the remaining groups. Further experiments will show if this is justified.

Assuming, then, that we are dealing with a linear function, we shall wish to know the probable standard error of an assay. The most favourable conditions under which a substance of unknown potency can be compared with a standard are that equal numbers of responses are obtained to each and the mean responses are the same. Then the standard error of the log ratio of potencies, s_M, will be

$$s_M = \frac{s}{b}\sqrt{\frac{2}{n_s}} = \sqrt{\frac{2Ve}{n_s b^2}}$$

where n_s is the number of responses to each substance. This follows from the fact that the variance of the difference between two means is the sum of their separate variances, and that we have assigned to each $\bar{Y}$ a variance of $\frac{Ve}{n_s}$.

The number of observations with each substance required to attain a given level of accuracy is thus at least:

$$n_s = \frac{2Ve}{b^2 VM}$$

In the assay of thyrotrophin examined in Chapters 6, 7 and 8, if we desired the estimate of potency to be ascertained within about ±20% of its true value in 95% of cases, then, s_M must be about 0·04, for antilog ±0·04t is approximately 0·8 to 1·2 if n is greater than about 20. If we correct for covariance, Ve is 42·1 and

$$n_s = \frac{2 \times 42{\cdot}1}{(40{\cdot}11)^2 \times (0{\cdot}04)^2} = 32{\cdot}7$$

We shall thus require at least 33 animals per substance, however they are grouped. Twice as many animals would be required if

we did not plan to correct by covariance, as the value of *Ve* for such a test is 84·0.

The number of observations we may require to attain a certain standard of accuracy may in practice exceed the predicted minimal number appreciably, by up to even 100% if our dosage groups are not very happily chosen.

9.3. Combining observations

If a dose-response line is calculated from the results of a well-planned experiment, it will necessarily have been determined from simultaneous observations. It is, however, a frequent and often necessary custom of research workers to conduct pilot tests with a new substance, and to put a few animals on one dose at a time to see how the responses are shaping. While a certain amount of such work may be needed to find the range of doses to which graded responses are to be expected, it is not wise to combine these haphazard observations into a single set and to attempt to treat them statistically. They should form the basis for a well-set-up test, to be conducted as soon as the investigator feels that he knows enough about the responses to the substance not to waste time and material by giving the wrong range of doses.

If observations made at various times are combined, either by adding to groups of observations on previously tested doses or by adding groups on new doses, the form and position of the dose-response line may be erroneously estimated unless the new data are added in a planned manner, with restrictions in design to allow for the segregation of a mean square attributable to time-to-time variation in response.

Temporal variation in response is frequently met in all types of assay, and is a fruitful source of annoyance, as an admirably planned experiment may go astray simply because the position of the line relating dose to response has shifted beyond the predicted limits and some of the responses are useless. This cannot be helped, unless it is due to laxity in controlling the experimental conditions, and the worker has no option but to repeat his work if the amount of information salvaged from the previous test is insufficient.

9.4. Missing items

In tests such as the Latin square, where the restricted randomisation requires an equal number of observations per group as an integral part of the design, the death of an animal or the dropping of

an observation for other causes upsets the analysis of the test and an estimate of it must be made. One missing item may be supplied in a test designed to segregate one source of possible variation other than regression and error, such as in Table 6.1, by the following formula:

$$Y' = \frac{dA + rB - T}{(d-1)(r-1)}$$

where Y' = the missing item,
d = the number of dosage groups,
r = the number of arrays in the restriction in design (i.e. the number of litters),
A = the sum of items receiving the same dose as Y',
B = the sum of items in the same array as Y',
T = the sum of all known items.

If the weight of the first uterus in litter 8 of Table 6.1 had been missing we should estimate it as

$$Y' = \frac{3 \times 452 + 8 \times 220 - 2{,}133}{2 \times 7}$$
$$= 70{\cdot}2.$$

One missing item in the doubly restricted Latin square may be replaced by:

$$Y'' = \frac{N(A + R + C) - 2T}{(N-1)(N-2)}$$

where N = the number of dosage groups (=number of rows or columns),
A = the sum of items receiving the same dose as Y',
R = the sum of items in the same row as Y',
C = the sum of items in the same column as Y',
T = the sum of all known items.

If more than one item is missing, but it seems worth while to go on with the analysis, a reiterative method may be used. First enter a reasonable value for Y' and use it for determining Y'', the second missing item. Then insert the value found for Y'' and calculate Y' by leaving it out, and so on until successive values of Y' and Y'' are practically identical. With three or more items the process is the same, but correspondingly tedious. The degrees of freedom available for estimating the variance due to random

sampling must be reduced by one for every missing item that has been estimated.

In general, the method minimises the sum of squares for random sampling. If we actually insert the algebraic symbol Y' and work out the analysis of variance, we shall get an expression:

$$Sy_e^2 = A - 2BY' + CY'^2$$

This is a minimum when $Y' = \frac{B}{C}$, and the sum of squares for error is then $A - \frac{B^2}{C}$, with one degree of freedom dropped.

9.5. Limitations of the analysis of variance

If the variance is not independent of response the analysis of variance loses theoretical validity. A fundamental condition of the F-test is that the two mean squares which are compared shall be independent. Frequently the variance of a mean response is correlated with the response—often the percentage standard error, or coefficient of variation, is relatively constant. When this is so, logarithms may be substituted for the original data before analysis, but if the data fitted by the original log dose-response line did not depart significantly from linearity, it is more than likely that the new line will not fit them as well, and we shall be faced with a new difficulty. It is fortunate that slight degrees of correlation between mean and variance and quite large departures from normality do not seriously affect the F-test, so that this type of difficulty, although it must be guarded against, is not too often insurmountable. If we decide to ignore a small correlation between variance and mean we shall be well advised to require a higher degree of significance in the F-test, say a P of 0·01 instead of 0·05, as we are really misusing the test and must play for safety. However, in a balanced test, when the mean responses to equally distributed groups on the standard and unknown are nearly the same, high correlations may often be safely ignored. The question is discussed further in Chapter 19.5.

9.6. Test of homogeneity of variance

When any doubt arises about the homogeneity of the variance of several groups of observations, Bartlett's test should be applied. Sample calculations for such a test are given in Table 9.1, in which

five groups of varying numbers of observations are tested for inequality of variance.

An index of dispersion, χ^2, which is described on page 134, is calculated from the approximate formula:

$$\chi^2 = \log_e 10[\log \bar{V}.S(n_p - 1) - S(n_p - 1) \log V_p],$$

where V_p is a variance based on $(n_p - 1)$ degrees of freedom, and $\bar{V}$ is the pooled estimate of variance from k samples, each yielding a separate estimate, V_p. The value of $\log_e 10$ is 2·3026.

TABLE 9.1

TEST OF HOMOGENEITY OF VARIANCE

Degrees of freedom $(n_p - 1)$	Sy^2	Variance (V_p)	$\frac{1}{n_p - 1}$	$\log_{10} V_p$	$(n_p - 1) \log V_p$
9	45·918	5·102	0·11111	0·70774	6·3697
7	16·387	2·341	0·14286	0·36941	2·5859
10	88·090	8·809	0·10000	0·94493	9·4493
5	20·365	4·073	0·20000	0·60991	3·0496
8	72·656	9·082	0·12500	0·95819	7·6655
Sums 39	243·416		0·67897		29·1200

In Table 9.1, we list the degrees of freedom in each sample, $(n_p - 1)$; the sum of squares, Sy^2; the mean square, $\bar{V}_p$; the reciprocal of $(n_p - 1)$; the log of V_p, and then multiply the last two together. Then, in the present example:

$$\bar{V} = \frac{SSy^2}{S(n_p - 1)} = \frac{243{\cdot}416}{39} = 6{\cdot}241$$

$$\log \bar{V}.S(n_p - 1) = 0{\cdot}79525 \times 39 = 31{\cdot}0148$$

$$\chi^2 = 2{\cdot}3026\ [31{\cdot}0148 - 29{\cdot}1200]$$

$$= 4{\cdot}363$$

This estimate of χ^2 should be corrected by a factor, C, such that:

$$C = 1 + \frac{1}{3(k-1)}\left[S\frac{1}{n_p - 1} - \frac{1}{S(n_p - 1)}\right]$$

unless, as in the present example, it is not significant as it stands,

when the correction is superfluous. Applying the correction for the purpose of illustration, we find:

$$C=1+\frac{1}{3\times 4}\left[0{\cdot}67897-\frac{1}{39}\right]$$

$$=1{\cdot}0544$$

Then $$\chi_c^2=\frac{\chi^2}{C}$$

$$=\frac{4{\cdot}363}{1{\cdot}0544}=4{\cdot}14 \text{ approx.}$$

This is tested with the Table of χ^2 (page 132), entering the Table with $k-1$ degrees of freedom, 4 d.f. in the present example. The value of χ^2 for 4 d.f. would have to exceed 9·488 to indicate significant heterogeneity at the 5% level.

CHAPTER 10

THE ESTIMATION OF RELATIVE POTENCY

10.1. Comparison of the action of two substances

When the potency of a preparation of unknown strength is to be compared with that of a standard preparation, we shall, in an adequately designed assay, have protocols from which two separate dose-response lines may be calculated. If these two lines depart significantly from parallelism, the relative potency of the substances depends on the particular dosage level at which the comparison is made. It is therefore possible validly to compare an unknown with a standard preparation only when the two dose-response lines are substantially parallel. Tests for significant departures from parallelism will be explained later in this chapter.

When the two dose-response lines do not differ significantly in slope, they are to be regarded as two separate estimates of a common slope relating response to the dose of both preparations. We therefore pool the information from both samples and calculate one value of b in the equation

$$E=a+bX,$$

where, it will be remembered, E is the estimate of response. The relative potency of the two preparations then depends on the difference between the two mean doses and the difference between the two mean responses to all doses. If M is the log ratio of the potency of the unknown to that of the standard,

$$M=\bar{X}_s-\bar{X}_u+\frac{\bar{Y}_u-\bar{Y}_s}{b}$$

where $\bar{X}_u$ is the mean log dose of the unknown and $\bar{X}_s$ is the mean log dose of the standard, $\bar{Y}_u$ the mean response to the unknown and $\bar{Y}_s$ the mean response to the standard.

10.2. Use of the method in practice

We now examine the determination of relative potency in the case of a simply planned assay of the potency of a preparation of

insulin. This assay involves only the restrictions that there shall be equal numbers of observations per dose, equal numbers of doses per substance and equal logarithmic spacing of doses.

In assaying the potency of insulin by measuring the fall in blood sugar of rabbits after injection of the drug, groups of animals are taken at random and their blood sugars measured by chemical estimation. The drug is then injected and the fall in blood sugar is measured over a period of several hours by taking successive samples of blood, and the mean percentage fall over the period of examination is used as the response. This is a rather elaborate

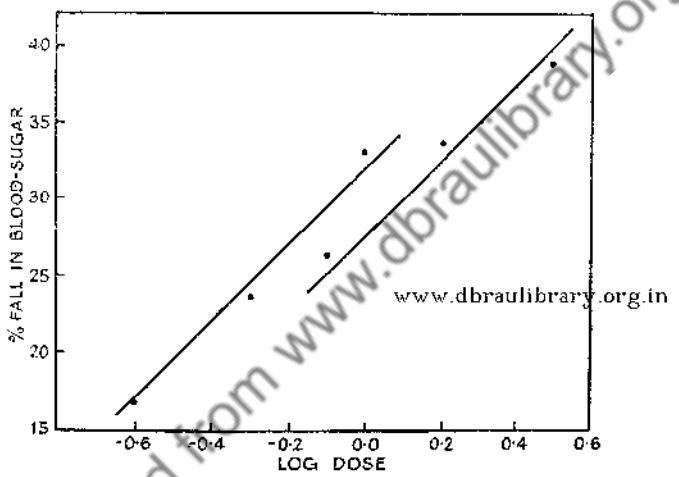


FIG. 10.1. A 2×3 assay of the potency of a sample of insulin. The standard is measured in units and the unknown in mg. (Table 10.1.) A common slope has been fitted to the data.

mode of expression and was introduced because the percentage fall in blood sugar is more constant than the absolute fall after any given dose. Covariance analysis, in which the initial level of blood sugar would be used as the concomitant variable, would make this rather arbitrary step unnecessary.

A series of such responses is shown in Table 10.1, which is adapted from Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939. Six groups, each of eight rabbits, are involved, three of them receiving three different doses of the standard and the other three receiving three different doses of the unknown. The assay is illustrated by Figure 10.1. The ratio of successive doses of both the standard and unknown is two. In Table 10.1 we have summed

the total responses to each dose and the total responses to each substance and also recorded the grand total.

TABLE 10.1

PROTOCOLS OF AN ASSAY OF THE POTENCY OF A PREPARATION OF INSULIN

(Adapted from Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

	Standard (units) Dose			Unknown (mg.) Dose		
	0·25	0·50	1·0	0·8	1·6	3·2
Responses: (% fall in blood sugar)	11·2	16·5	32·7	19·8	37·7	45·4
	21·2	23·2	14·0	21·7	40·7	28·6
	18·7	25·6	28·9	26·1	29·3	50·4
	2·8	12·7	40·2	32·2	48·1	47·7
	27·2	39·8	35·1	28·5	45·6	50·0
	25·1	28·4	36·2	20·2	35·3	12·4
	25·8	40·0	37·8	35·7	14·2	39·0
	2·2	2·4	39·4	26·1	7·9	38·1
Totals, T_p	134·2	188·6	264·3	210·3	258·8	311·6
Grand totals	587·1 ($=T_s$)			780·7 ($=T_u$)		
	$T=1{,}367{\cdot}8$					

Sums of squares

$$\bar{y}_s^2+\bar{y}_u^2=780{\cdot}85$$
$$S\bar{y}_p^2=2{,}489{\cdot}92$$
$$Sy^2=7{,}776{\cdot}71$$

It is sometimes worth while to make a preliminary analysis of variance on such material, the results of which are shown in Table 10.2. The sums of squares of the difference between the mean response to each substance (all groups combined) and the general mean response, which is:

$$\bar{y}_s^2+\bar{y}_u^2$$

is determined in the usual way and is found to be 780·85, and with it is associated a single degree of freedom. The sums of squares of the differences between the group means and the general mean, $S\bar{y}_p^2$, is found to be 2,489·92, and with it are associated five degrees of freedom, the sum of squares associated with one of which has already been determined, leaving four degrees of freedom for the sum of squares between doses of the same substance, this sum of squares being:

$$2{,}489{\cdot}92-780{\cdot}85=1{,}709{\cdot}07$$

The sum of squares of all deviations from the general mean $\bar{Y}$, Sy^2, is 7,776·71; from this we subtract 2,489·92 in order to obtain the sum of squares for random sampling. This completes the preliminary analysis of variance shown in the Table. Each mean square is divided by the mean square for random sampling, with which are associated 42 degrees of freedom (6×7 degrees of freedom, since there are seven independent comparisons within each dosage group) and it is seen that the values of F obtained are significant but not highly significant, P being less than 0·05 but greater than 0·01 in each instance. When we proceed to examine the results of this assay it will be well to bear in mind that we may be dealing with a low level of significance and of accuracy, but we are not sure of this until the further analysis is made.

TABLE 10.2

PRELIMINARY ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 10.1

Source of variation	Degrees of freedom	Sum of squares	Mean square	F	P
Between samples	1	780·85	780·9	6·2	<0·05
Between doses of the same substance	4	1,709·07	427·3	3·4	<0·05
Random sampling	42	5,286·79	125·9		
Total	47	7,776·71			

10.3. Factorial coefficients

In balanced designs of the type we are examining, where equal numbers of doses of the standard and unknown at equally spaced logarithmic intervals are given to groups of equal numbers of animals, we simplify the analysis of the dose-response lines by the use of *factorial coefficients*. Factorial coefficients are a logical development of the polynomial coefficients we examined in dealing with a single log dose-response line. Most comparisons of the potency of a standard and unknown can be made by administering two, three or four doses of each substance. When we have examined the dose-response relationship for the standard prior to conducting assays—as we should have done—it will rarely be necessary to exceed four doses of a substance.

Denoting the 1–4 doses of the standard by S_1–S_4 and those of the unknown by U_1–U_4, we may tabulate the factorial coefficients to be used in the isolation of individual effects, as in Table 10.3. These coefficients are the factors by which we shall multiply the total of responses on each dosage group of the two preparations.

TABLE 10.3

FACTORIAL COEFFICIENTS FOR THE ISOLATION OF INDIVIDUAL EFFECTS WHEN TWO, THREE OR FOUR DOSES OF THE STANDARD AND UNKNOWN ARE ADMINISTERED

Source of variation	Factorial coefficients (k) for dose:																	
	S_1	S_2	U_1	U_2	S_1	S_2	S_3	U_1	U_2	U_3	S_1	S_2	S_3	S_4	U_1	U_2	U_3	U_4
1. Differences between samples	−1	−1	+1	+1	−1	−1	−1	+1	+1	+1	−1	−1	−1	−1	+1	+1	+1	+1
2. Linear regression	−1	+1	−1	+1	−1	0	+1	−1	0	+1	−3	−1	+1	+3	−3	−1	+1	+3
3. Departure from parallelism	+1	−1	−1	+1	+1	0	−1	−1	0	+1	+3	+1	−1	−3	−3	−1	+1	+3
4. Curvature of combined curve	—	—	—	—	+1	−2	+1	+1	−2	+1	+1	−1	−1	+1	+1	−1	−1	+1
5. Opposed curvature of separate curves	—	—	—	—	−1	+2	−1	+1	−2	+1	−1	+1	+1	−1	+1	−1	−1	+1
6. Double curvature of combined curve	—	—	—	—	—	—	—	—	—	—	−1	+3	−3	+1	−1	+3	−3	+1
7. Opposed double curvature	—	—	—	—	—	—	—	—	—	—	+1	−3	+3	−1	−1	+3	−3	+1

When two doses of the unknown and two doses of the standard are administered there are three degrees of freedom and thus three independent comparisons which may be made by factorial analysis. The three sources of variation concerned are those due to differences between the response to the two samples, to linear regression and to departure of the two dose-response lines from parallelism. In this example, it is easy to see the logic behind the application of the factorial coefficients; the difference in response between samples is clearly T_u–T_s, the linear regression coefficient is $\frac{\bar{Y}_2-\bar{Y}_1}{\bar{X}_2-\bar{X}_1}$ and the departure from parallelism is the difference between the two separate estimates of (Y_2-Y_1) for the standard and unknown.

When there are three doses of each substance, there are five degrees of freedom and in addition to examining the three sources of variation already enumerated, we may examine the curvature of the combined curve and the opposed curvature of the two separate curves, both of which should be insignificant in a satisfactory assay. When there are four doses of each substance, two more sources of variation may be examined, namely, those due to double curvature of the combined and separate curves. Thus, by the use of factorial coefficients we are able to extract all the information relevant to the dose-response lines, considered separately and together. The application of these coefficients now follows in the example under discussion.

10.4. Factorial analysis of the dose-response line

Table 10.4 shows the analysis for the data of Table 10.1. We read off the relevant factorial coefficients from the central columns of Table 10.3 and at the foot of each column representing a dosage group we note the totals of the responses in that group (T_p). These are the totals from Table 10.1. The factorial coefficients, which we will denote by k, are squared in each row and the total of Sk^2 multiplied by the number of observations per group, $=n_pSk^2$, is written in a column under the heading "divisor." Each total is then multiplied by the corresponding coefficient in each row, to form the quantity kT_p, and the sum of these quantities in each row, which is the sum of products, is written down in the next column. This sum is denoted by SkT_p. In forming it, the algebraic sign of each coefficient must be taken into consideration. The variance attributable to each source of variation is the square of the sum of products divided by the divisor, and with it is associated a single

degree of freedom. The sum of squares for random sampling is the error sum of squares and the mean square for error, in this case 125·9, is written at the foot of the variance column. The importance of each source of variation is then examined as usual by the F-test, entering the table of F in our present example with one and 42 degrees of freedom for the two groups respectively.

TABLE 10.4

FACTORIAL ANALYSIS OF THE DOSE-RESPONSE LINE FOR THE DATA OF TABLE 10.1

Source of variation	Factorial coefficients (k) for dose						Divisor $n_p Sk^2$	Sum of products $Sk.T_p$	Variance	
	S_1	S_2	S_3	U_1	U_2	U_3				
Differences between samples	−1	−1	−1	+1	+1	+1	48	193·6	780·9	(D^2)
Linear regression	−1	0	+1	−1	0	+1	32	231·4	1,673·3	(B^2)
Departure from parallelism	+1	0	−1	−1	0	+1	32	−28·8	25·9	
Curvature of combined curve	+1	−2	+1	+1	−2	+1	96	25·6	6·8	
Opposed curvature of separate curves	−1	+2	−1	+1	−2	+1	96	−17·0	3·0	
Totals, T_p	134·2	188·6	264·3	210·3	258·8	311·6			Error 125·9	

The mean square for the difference between samples is the same as in Table 10.2 and in fact we need not have determined the values separately in the previous Table. We denote this mean square by the symbol D^2 and its F value shows whether the potency of the actual doses of the unknown which were administered differed from those of the standard. We have already seen that in this assay they did. The variance attributable to linear regression, which we denote by B^2, measures the average increase in response due to equivalent increases in the doses of standard and the unknown, as determined by the combined slope. B^2 must be significantly greater than the error variance or the assay is not valid, since the slope does not differ significantly from zero. Our present B^2 gives an F of 13·3 and shows that the value of the slope is in fact highly significant—a point which was not apparent from the preliminary analysis of variance in Table 10.2. This high significance of B is found because the variances attributable to the

various departures from linearity and parallelism are particularly small.

10.5. Computation of the log ratio of potencies, using factorial coefficients

The equation given for M in Section 10.1 reduces, when factorial coefficients are used, to:

$$M=\bar{X}_s-\bar{X}_u+\frac{kID}{B}$$

where $k=1, \sqrt{8/3}$ and $\sqrt{5}$ for assays with two, three and four doses of each substance respectively, I is the interval in logarithms between successive doses, and D and B are the square roots of D^2 and B^2, the first two variances in Tables 10.3 and 10.4. For use in this equation, D and B must both be given the same sign as the sum of products from which they are computed. The terms k and I convert the log potency from the answer given by factorial coefficients back to normal logarithms.

In our present example we measured the doses of the standard in units, since we were using International Standard insulin, but we administered our unknown in mgm. We shall, of course, wish to find the potency of the unknown in terms of the number of units it contains per mgm. For this purpose it is merely necessary to assume that the standard and the unknown are of equal potency and that both are administered in the same dosage units and thus:

$$\bar{X}_s=\bar{X}_u$$

and the equation reduces to:

$$M=\frac{kID}{B}.$$

The antilog of M is the number of units of the standard required to give the same response as one "assumed unit" of the unknown. The highest dose of the standard was one unit; our assumed unit for the unknown is thus 3·2 mgm. Substituting in the equation, we find that:

$$M=\frac{\sqrt{8/3}\times 0{\cdot}30103\times\sqrt{780{\cdot}9}}{\sqrt{1{,}673{\cdot}3}}=0{\cdot}3358.$$

Thus the logarithm of the potency of 3·2 mgm. of the unknown is 0·3358, the antilog of which is 2·167. Hence, 3·2 mgm. contains 2·167 units, 1 mgm. contains 0·677 units of insulin.

10.6. The standard error of M

The standard error of M, s_M, is approximately given by the formula:

$$s_M = \frac{skI\sqrt{B^2 + D^2}}{B^2}$$

where s is the root mean square for error from the analysis of variance and all other terms have the same significance as in the equations for determining potency. We use s_M in conjunction with a table of t, with n, the number of degrees of freedom, equal to the number of degrees of freedom for experimental error in the analysis of variance. The value of t for any required degree of significance, usually for $P=0{\cdot}05$ or $0{\cdot}01$, is read from the table and multiplied by s_M; then the potency of the unknown preparation has been determined within the limits of antilog $(M+ts_M)$ and antilog $(M-ts_M)$. If we wish to know the percentage accuracy of the determination, we give M the value 2.

It will be noted that the logarithmic transformation has the result that the upper limit assigned to the value of the potency at any given level of significance is further from the most probable figure, given by antilog M, than is the lower limit. In our example:

$$s_M = \sqrt{125{\cdot}9} \times \sqrt{8/3} \times 0{\cdot}30103 \times \frac{\sqrt{2{,}454{\cdot}2}}{1{,}673{\cdot}3} = 0{\cdot}1673.$$

The value of t for $P=0{\cdot}05$ with 42 degrees of freedom is $2{\cdot}021$ and the limits of error at this level of significance are therefore the antilogs of $0{\cdot}3358 - 2{\cdot}021 \times 0{\cdot}1673$ and $0{\cdot}3358 + 2{\cdot}021 \times 0{\cdot}1673$, or $\bar{1}{\cdot}9977$ and $0{\cdot}6739$. These limits are therefore $0{\cdot}995$ and $4{\cdot}720$ for the relative potency on the assumption of equality of units. We divide these limits by $3{\cdot}2$ to obtain the number of units per mgm. of the unknown, which has therefore been determined within the limits $0{\cdot}311$ and $1{\cdot}475$. These limits are very wide, for we have determined the potency within a percentage error of $45{\cdot}9$ to $217{\cdot}8\%$ at the level of significance $P=0{\cdot}05$, and thus our value for antilog M could not with assurance be said to differ significantly from as much as half or twice its value. In the next chapter it will be seen that even these limits are not as wide as exact calculation reveals, since they are approximate limits, derived from a formula which closely approaches the exact formula only if the slope of the dose-response line has been determined with little error.

10.7. The four-point assay

An assay in which only two dose-levels of each substance are employed, usually called a "four-point assay," cannot supply evidence about the nature of the dose-response line. It should not be used unless it is already known that the log dose-response line is either straight or can be represented by a second-degree equation —i.e. it exhibits simple curvature. In the latter instance, it is a surprising fact that, with a balanced design, the relative potency is correctly estimated by the four-point assay although *not* by assays using more than four dosage groups in all (Gridgeman, N. T., *Biochem. J.*, **37**, 127, 1943; Wood, E. C., *Nature*, **153**, 680, 1944).

CHAPTER 11

THE ESTIMATION OF RELATIVE POTENCY WITH UNBALANCED DOSAGE GROUPS

11.1. A badly planned test

The relatively simple procedure available for the estimation of relative potency and its errors when a balanced design has been used in the test is a forceful argument in favour of the adoption of such designs. The addition of various restrictions in design, possible only in balanced designs, adds but little to the computational procedure and may reduce error very considerably.

We shall now consider the computational procedure which must be applied when the doses of the standard and unknown preparations are unequal in number and do not bear the same relation to one another, while the groups do not contain equal numbers of observations. This is the worst that can happen, and the reader will appreciate that it is a situation to be avoided, if possible, at almost any cost. Not only is the amount of information which may be extracted from the material much reduced, but the arithmetic becomes tedious.

TABLE 11.1

PROTOCOLS OF AN ASSAY OF OESTROGENIC HORMONE, WITH UNEQUAL GROUPS AND SPACING OF DOSES.

	Standard			Unknown	
Dose (μg.)	0·2	0·3	0·4	1·0	2·5
Responses (weight of uteri)	54	59	152	61	102
	49	85	71	74	73
	51	143	112	51	112
	81	60	58	60	130
	63	74	102	83	105
	126	72	—	83	118
	—	103	—	—	105
	—	110	—	—	131
Totals, T_p	424	706	495	412	876
Means, $\bar{Y}_p$	70·7	88·3	99·0	68·7	109·5
n_p	6	8	5	6	8

Table 11.1 gives the basic data relating to an assay of oestrogenic hormone, in which three doses of 0·2, 0·3 and 0·4 μg. of the standard preparation were given to three groups of unequal numbers of rats and two doses of 1 and 2·5 μg. of the unknown were given to two further unequal groups. The assay is shown in Figure 11.1. The procedure was the same as that described in Chapter 6 for the test using the response of rat uteri to injected oestrone. In Table 11.1 we list the total of each group, the means of each group, and it is useful to jot down the value of n_p itself below these figures. The means were used in the computations, but it should be noted that for full accuracy in working they should be taken to one or two more decimal places than has been done in this example. The

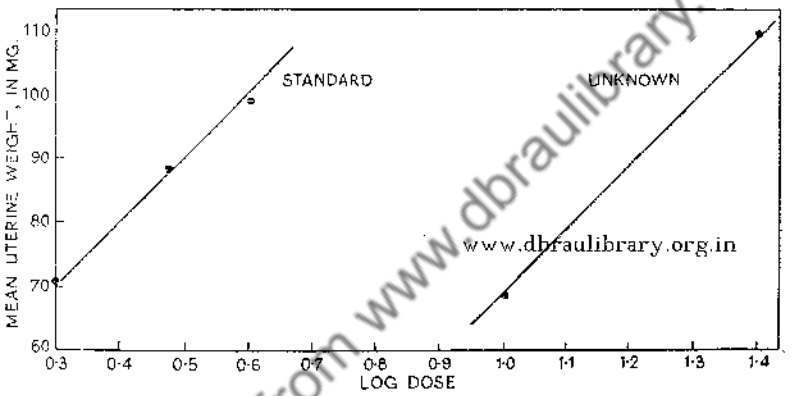


FIG. 11.1. An assay of oestrogenic hormone, with unbalanced design (Table 11.1). A common slope has been fitted to the data. To avoid negative numbers, each dose has been multiplied by 10.

effect in the present instance is trivial. It is, however, always safer to work in totals where possible. In Chapter 4.6 we saw how to calculate weighted means and a regression line when groups were not equal, and this procedure must be followed here. If there is any doubt about the significance of the slope of the combined dose-response curve, a point which can often be settled by inspection of the data, it is worth while to save involved computation by performing a preliminary analysis of variance, as in Table 11.2. On the other hand, if the investigator is reasonably sure that a valid assay has been performed, it is quicker to run the whole analysis together, as in Table 11.3. The determination of the separate quantities which we need to establish in composing Table 11.2 will therefore be described in the general

calculations based on Table 11.3 When the relevant data have been extracted, we shall therefore refer back to 11.2 and point out how the required quantities could have been filled in.

TABLE 11.2

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 11.1

Source of variation	Degrees of freedom	Sum of squares	Mean square	F	P
Between samples	1	337·8	337·8	0·50	>0·05
Between doses of the same substance	3	8,071·7	2,690·1	4·00	<0·05
Random sampling	28	18,845·0	673·0	—	—
Total	32	27,254·5	—	—	—

TABLE 11.3

CALCULATION OF REGRESSION FOR THE DATA OF TABLE 11.1

	log (dose ×10) $\bar{X}_p$	Mean response $\bar{Y}_p$	No. of rats n_p	$n_p\bar{X}_p$	$n_p\bar{X}_p^2$	$n_p\bar{Y}_p$	$n_p\bar{X}_p\bar{Y}_p$	$n_p\bar{Y}_p^2$
Standard	0·301	70·7	6	1·806	0·54361	424	127·62	29,976·8
	0·477	88·3	8	3·816	1·82023	706	336·76	62,339·8
	0·602	99·0	5	3·010	1·81202	495	297·99	49,005·0
		Totals	19	8·632	4·17586	1,625	762·37	141,321·6
Unknown	1·000	68·7	6	6·000	6·00000	412	412·00	28,304·4
	1·398	109·5	8	11·184	15·63523	876	1,224·65	95,922·0
		Totals	14	17·184	21·63523	1,288	1,636·65	124,226·4
	Grand Totals		33	—	—	2,913	—	265,548·0

11.2. Combined calculation for analysis of variance and regression

The calculations which are required for this purpose are exemplified in Table 11.3. We list for the standard and unknown, first, $\bar{X}_p$, the logarithm of 10 times the dose (to avoid negative logarithms we multiply the dose by 10), next, Y_p, the mean response in each group, and n_p, the number of rats in each group. From the third column onwards we also require separately for the standard and unknown the sums of each column and the grand totals for both standard and unknown. The last five columns of the Table give the quantities required in calculating weighted

means, sums of squares and sums of products. From them we determine the following quantities:

For the standard, $\bar{X}_s = \frac{Sn_p\bar{X}_p}{Sn_p} = \frac{8{\cdot}632}{19} = 0{\cdot}45432$

For the unknown, $\bar{X}_u = \frac{Sn_p\bar{X}_p}{Sn_p} = \frac{17{\cdot}184}{14} = 1{\cdot}22743$

For the standard, $\bar{Y}_s = \frac{Sn_p\bar{Y}_p}{Sn_p} = \frac{1{,}625}{19} = 85{\cdot}526$

For the unknown, $\bar{Y}_u = \frac{Sn_p\bar{Y}_p}{Sn_p} = \frac{1{,}288}{14} = 92{\cdot}000$

For both standard and unknown together,

$$\bar{Y} = \frac{Sn_p\bar{Y}_p}{Sn_p} = \frac{2{,}913}{33} = 88{\cdot}273$$

In the above, SY_s and SY_u may be read for the respective quantities $Sn_p\bar{Y}_p$.

For the standard:

$$Sn_p\bar{y}_p^2 = Sn_p\bar{Y}_p^2 - \frac{T_s^2}{n_s} = 141{,}321{\cdot}6 - 138{,}980{\cdot}3 = 2{,}341{\cdot}3$$

$$Sy_s^2 = SY_s^2 - \frac{T_s^2}{n_s} = 156{,}905{\cdot}0 - 138{,}980{\cdot}3 = 17{,}924{\cdot}7$$

$$Sn_p\bar{x}_p\bar{y}_p = Sn_p\bar{X}_p\bar{Y}_p - \bar{X}_sT_s = 762{\cdot}37 - 738{\cdot}27 = 24{\cdot}10$$

$$Sn_p\bar{x}_p^2 = Sn_p\bar{X}_p^2 - \left(\frac{Sn_p\bar{X}_p}{n_s}\right)^2 = 4{\cdot}17586 - 3{\cdot}92165 = 0{\cdot}25421$$

where Sy_s^2 is to be determined from the sums of squares of all the individual observations on the standard.

For the unknown, by analogy with the above:

$$Sn_p\bar{y}_p^2 = 5{,}730{\cdot}4$$
$$Sy_u^2 = 8{,}992{\cdot}0$$
$$Sn_p\bar{x}_p\bar{y}_p = 56{\cdot}56$$
$$Sn_p\bar{x}_p^2 = 0{\cdot}54310$$

For both standard and unknown:

$$Sy^2 = SY^2 - \frac{T^2}{n} = 27{,}254{\cdot}5$$

The analysis of variance in Table 11.2 is constructed as follows. The variance associated with differences between samples, with which is associated one degree of freedom, is:

$$Sy^2 - Sy_s^2 - Sy_u^2 = 337{\cdot}8$$

This may be checked by determining the quantity:

$$n_s\bar{y}_s^2+n_u\bar{y}_u^2=n_s(\bar{Y}_s-\bar{Y})^2+n_u(\bar{Y}_u-\bar{Y})^2=337{\cdot}8$$

The variation between doses of the same substance is, for both the standard and unknown:

$$SSn_py_p^2=2{,}341{\cdot}3+5{,}730{\cdot}4=8{,}071{\cdot}7$$

with which are associated three degrees of freedom.

The sum of squares attributable to random sampling is determined by subtraction of these two quantities from Sy^2 and is:

$$27{,}254{\cdot}5-8{,}071{\cdot}7-337{\cdot}8=18{,}845{\cdot}0$$

These mean squares, compared with the mean square for random sampling by the F-test, indicate that while there is in fact no significant difference between the mean responses to samples, the regression of response on log dose, as measured by the mean square for differences between doses on the same substance, is significant, but not highly significant. There are two further sources of variation which may be examined, namely, departure of the three points determined for the standard substance from a linear dose-response relationship, and the departure of the two separate dose-response lines from parallelism. The first of these departures may be tested, as explained in Chapter 4.6, using only the data relevant to the standard, since the unknown contributes no information on this point. The deviation from regression, with which in this case is associated one degree of freedom, is compared with the error mean square as in Table 11.4. It will be noted that neither it nor the regression for the standard alone is significant. It is only the combined evidence of both samples that enables us to postulate a significant regression of response on dose. The second source of variation cannot be examined until we have learnt

TABLE 11.4

ANALYSIS OF VARIANCE FOR THE STANDARD ALONE FROM THE DATA OF TABLE 11.1

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square	F	P
A. Linear regression	$\frac{(Sn_p\bar{x}_p\bar{y}_p)^2}{Sn_p\bar{x}_p^2}$	1	2,280·8	2,280·8	2·34	>0·05
B. Deviation from regression	$Sn_p\bar{y}_p^2-A$	1	60·5	60·5	0·062	>0·05
C. Random sampling	$Sy^2-(A+B)$	16	15,583·4	974·0	—	—
Total	Sy^2	18	17,924·7	—	—	—

how to calculate the variance of the slope. Neither will be found to be significant. The analysis of variance having indicated that a valid assay is possible, although not with as high a degree of significance as we might desire, we may proceed to calculate the combined slope and its errors.

11.3. Calculation of the slope and its standard error

The combined slope for the two substances, b, is given by the equation:

$$b=\frac{SSn_p\bar{x}_p\bar{y}_p}{SSn_p\bar{x}_p{}^2}$$

where SS means "the sum of the sums of . . ." since we add together the two separate totals for the standard and unknown as described above. We also determine the slopes separately because we shall require to examine them separately. Each separate slope is given by the equation:

$$b=\frac{Sn_p\bar{x}_p\bar{y}_p}{Sn_p\bar{x}_p{}^2}$$

The combined slope is thus:

$$\frac{24{\cdot}10+56{\cdot}56}{0{\cdot}25421+0{\cdot}54310}=101{\cdot}17$$

The two separate slopes are:

$$b_s=94{\cdot}80 \text{ and } b_u=104{\cdot}14$$

for the standard and unknown respectively.

The variance of the slope is:

$$Vb=\frac{Ve}{SSn_p\bar{x}_p{}^2}=\frac{18{,}845{\cdot}0}{28\times 0{\cdot}79731}=844{\cdot}18$$

whence $s_b=29{\cdot}05$.

It might be judged from the figures involved that the two separate estimates of the slope do not differ significantly, but we may if we wish test the point by determining a separate variance for each slope, that for the standard being given by:

$$Vb_s=\frac{Ve}{Sn_p\bar{x}_p{}^2}$$

where $Sn_p\bar{x}_p{}^2$ refers only to the doses of the standard. The corresponding variance for the slope of the unknown is similarly determined. The reader may care to do this as an exercise and will find no significant difference between the two slopes. The

variance of the difference between the slopes is given by the formula:

$$V(b_s - b_u) = Vb_s + Vb_u$$

and "Student's" t is $\frac{b_s - b_u}{\sqrt{V(b_s - b_u)}}$, with the same number of degrees of freedom as for determining Ve.

11.4. The estimation of relative potency

From the formula:

$$M = \bar{X}_s - \bar{X}_u - \frac{\bar{Y}_s - \bar{Y}_u}{b}$$

we determine relative potency as before.

$$M = -0{\cdot}77311 - \frac{-6{\cdot}474}{101{\cdot}17} = -0{\cdot}7091$$

Antilog $-0{\cdot}7091$, or $\bar{1}{\cdot}2909$, is $0{\cdot}1954$. The unknown is thus $0{\cdot}1954$ times as potent as the standard.

The approximate formula for the error of M is given by:

$$VM = \frac{Ve}{b^2}\left(\frac{1}{n_s} + \frac{1}{n_u}\right) + \frac{(\bar{Y}_s - \bar{Y}_u)^2 Vb}{b^4}$$

$$= \frac{673{\cdot}07}{101{\cdot}17^2}\left(\frac{1}{19} + \frac{1}{14}\right) + \frac{6{\cdot}474^2 \times 844{\cdot}18}{101{\cdot}17^4}$$

$$= 0{\cdot}008496,$$

whence $s_M = 0{\cdot}09217$.

The limits of error of our determination of potency at any given level of significance are therefore given by the antilogs of:

$$-0{\cdot}7091 \pm 0{\cdot}09217t$$

where t corresponding to any desired level of significance is read from the Table with 28 degrees of freedom. The value of t for $P = 0{\cdot}05$ is $2{\cdot}048$ and for $P = 0{\cdot}01$ is $2{\cdot}763$. Substituting at these two levels of significance, we find for $P = 0{\cdot}05$,

$$-0{\cdot}7091 \pm 0{\cdot}1888,$$

and for $P = 0{\cdot}01$,

$$-0{\cdot}7091 \pm 0{\cdot}2547$$

whence, taking antilogs, we obtain the limits $0{\cdot}1265$ to $0{\cdot}3018$ for $P = 0{\cdot}05$ and $0{\cdot}1087$ to $0{\cdot}3594$ for $P = 0{\cdot}01$.

11.5. Fiducial limits of error

It has been indicated that the limits given so far for the estimation of relative potency are approximate. The approximate formula is sufficiently accurate as long as significance of the slope, b, is high. However, when our estimate of the slope of the log dose-response line is not well determined, the approximate formula may give grossly misleading results. It is possible to calculate exact limits, usually called *fiducial limits*, in which the error of the determination of the slope is taken fully into account. The point is fully discussed by Irwin (*J. Hygiene*, **43**, 121, 1943). When the highest possible accuracy is required it is always worth while to calculate exact fiducial limits unless the slope is greater than about eight times its standard error.

The fiducial limits of the results of an assay require the calculation of the following quantities:

$$A = Ve\left(\frac{1}{n_s} + \frac{1}{n_u}\right) = s^2\left(\frac{1}{n_s} + \frac{1}{n_u}\right)$$

$$Vb = \frac{Ve}{SSn_p\bar{x}_p^2}$$

$$C = \frac{b^2}{b^2 - t^2 Vb} = \frac{B^2}{B^2 - s^2 t^2}$$

The limits are given by:

$$\bar{X}_s - \bar{X}_u - \frac{C(\bar{Y}_s - \bar{Y}_u)}{b} \pm \frac{t\sqrt{C}}{b}\left(A + \frac{C \cdot Vb(Y_s - Y_u)^2}{b^2}\right)^{\frac{1}{2}}$$

When factorial coefficients have been used in the calculation of the potency and dose-response relationship the limits are:

$$CM \pm \frac{2st\sqrt{C}}{b}\left(1 + \frac{D^2}{B^2 - s^2t^2}\right)^{\frac{1}{2}}$$

(Bliss, *Biometrics Bulletin*, **1**, 57, 1946) where n_pSk^2 is the number per group times the sum of the squares of the coefficients for the row determining the difference between samples in the factorial analysis.

For an odd number of dosage groups:

$$b = \frac{B}{I\sqrt{n_pSk^2}}$$

and for an even number of dosage groups:

$$b=\frac{2B}{I\sqrt{n_p Sk^2}}$$

where $n_p Sk^2$ is taken from the linear regression now in the factorial analysis.

It will be seen that these limits must be determined for each level of significance which it is desired to investigate, as the first part of each expression giving the limits is not identical with M, the log ratio of potencies. The range of fiducial limits does not have M as its central point. Thus, in addition to underestimating the width of the fiducial range, the approximate formula gives a biassed estimate of its position, unless $\bar{Y}_s=\bar{Y}_u$. The approximate formula equates C to unity, which is only the case if Vb is infinitely small, but C rapidly approaches unity as $\frac{b}{s_b}$ increases. When $\frac{b}{s_b}=8$, the calculation of approximate limits involves about a 5% error in the value of C at the level $P=0{\cdot}05$, and the fiducial limits differ very little from the approximate limits.

11.6. Calculation of fiducial limits in a practical example

The ratio $\frac{b}{s_b}$ in the example we have been examining in this chapter is only 3·5. Hence we may expect that the fiducial limits, particularly for high levels of significance, will differ considerably from the limits as determined by the approximate formula. We calculate:

$$A=673{\cdot}07\left(\frac{1}{19}+\frac{1}{14}\right)=83{\cdot}501$$

$$Vb=\frac{673{\cdot}07}{0{\cdot}79731}=844{\cdot}18$$

$$C=\frac{101{\cdot}17^2}{101{\cdot}17^2-2{\cdot}763^2\times 844{\cdot}18}=2{\cdot}700$$

$$\sqrt{C}=1{\cdot}643$$

where we have taken t for the 0·01 level of significance = 2·763. Substituting in the formula for fiducial limits, we find that these limits are given by $-0{\cdot}6003\pm 0{\cdot}4302$ and are thus $2{\cdot}9695-1{\cdot}8299$. Taking antilogs, the potency of the unknown is thus judged to fall within the limits 0·09322 to 0·6759. Since the estimate of the most probable potency (antilog M) was 0·1954, we calculate that

the potency has thus been determined within the limits 47·2% to 345·9% of its most probable value at the level of significance $P=0{\cdot}01$.

Our estimate from the approximate formula was 0·1087 to 0·3594, or 55·6% to 183·9%. This very large difference arises because the value of 3·5 for $\frac{b}{s_b}$ is unsatisfactory (also $\bar{Y}_s$ differed a lot from $\bar{Y}_u$), and serves as a warning that when the significance of the slope is not very high, fiducial limits must be calculated in the place of the approximate limits. At the level of probability where the slope loses significance and cannot be considered to differ from zero, the fiducial limits for the result are 0 and infinity, but the approximate formula still gives finite limits at this level. It is important, therefore, to calculate exact limits for a high level of significance, if that level of significance is not far exceeded by the level of significance of the value of the slope.

The approximate limits calculated for the assay discussed in Chapter 10 were, at the level of probability $P=0{\cdot}05$, 45·9% and 217·8%. If the reader cares, as an exercise in the calculation of fiducial limits, to compute them for this assay, he will find that the fiducial limits are 53·2% and 373·7%. Note that in this case not only was the range of error underestimated but that the lower limit of error given by the approximate formula was in fact below the true value, whereas the upper limit was considerably short of it. This fairly serious discrepancy between the fiducial and approximate limits at a probability level of 0·05 is again attributable to the fact that $\frac{b}{s_b}$ takes a low value (approximately 3·6). The two examples in Chapters 10 and 11 were in fact selected with a view to impressing from the start the need for the calculation of exact limits of error when the slope has not been determined with any very high degree of precision. In a great number of well-designed and well-conducted tests, such as that we are about to discuss in the next Chapter, there is no need for the calculation of limits other than those indicated by the approximate formula, since the slope will have been determined with considerable exactitude.

CHAPTER 12

A 2×4-POINT ASSAY WITH RESTRICTIONS IN DESIGN

12.1. Protocols of the assay

The majority of examples of adequately designed and adequately analysed assays is found in the published works of a very few authors. We owe in particular a considerable debt to the work of Dr. C. I. Bliss and his collaborators, from whose published material the following example is taken (Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939). It illustrates the treatment of an assay (a "2×4 assay") in which four doses of the standard and four doses of the unknown are given in a balanced design in which the variation attributable to differences between animals and the two successive days on which they were injected may be segregated.

The example compares the relative potency of International Standard insulin on two successive days and was made for the purpose of testing the action of the drug during the restricted food intake of the test, but the comparison of the action of the same compound on two successive days involves exactly the same statistical treatment as does the comparison of standard and an unknown substance on the same day. International Standard insulin administered on Day 1 may be regarded as the standard and the same compound administered on Day 2 as the unknown.

The test also exemplifies the comparison of the reactions of the different animals (rabbits) used in the "cross-over" technique. Each rabbit receives each of the four doses once on the first day and once on the second day, but within these restrictions the animals are assigned at random to doses. As described in Chapter 10, the test is conducted by the injection of a dose of insulin and the measurement of the percentage fall in blood sugar which it evokes. It is not possible to make four such measurements on one animal in one day, and Days 1 and 2 therefore refer only to the first and second of four different pairs of consecutive days on which each animal was used. Since the initial level of blood sugar was known to have an influence on the percentage fall, this level was also recorded for use in correction by covariance analysis.

The protocols of the test are given in Table 12.1. As often happens in practice, certain accidents during the course of the test had the result that a few rabbits were rejected from the experiment, and of the 16 rabbits originally used only 12 were retained for analysis. This disturbed the original balance of the experiment, so that the different doses were not equally represented on each day of the test and few substitutions were made to restore the approximate balance. These are indicated in the Table. In Bliss's view, since separate analysis showed that the apparent variation in sensitivity from one day to another could be accounted for largely by changes in initial blood sugar, the partial lack of balance between doses on any one day should be adjusted automatically in the correction by covariance for differences in initial blood sugar. These details need not worry us further, since the example is merely illustrative of the statistical methods to be employed on the assumption of complete balance.

12.2. The analysis, using covariance

In Table 12.1 the total reactions for each rabbit and the total reactions of all rabbits on each dosage level are tabulated, together with the corresponding totals for the initial levels of blood sugar measured in mgm./100 ml. $\bar{X}_p$ is the logarithm to base 10 of a hundred times the actual dose administered.

The data in Table 12.1 are now analysed, using factorial coefficients, as in Table 10.3, and covariance analysis to adjust for initial differences in blood sugar. The first stage in the analysis is shown in Table 12.2. In this Table we list the coefficients for treatment effects Nos. 1–7 of Table 10.3, and below each column of coefficients we write the totals, t_p and T_p, the total of initial blood sugars and percentage reduction in blood sugars at each dosage level on each day. The divisor, $n_p Sk^2$, is written against each row of factorial coefficients and the two sums of products for W and Y fall in the next two columns. Then follow three columns of variances and covariances, from which an adjusted Sy_a^2 for each treatment effect will be calculated.

The coefficients for adjusting Sy^2 shown at the bottom of the columns of variances and covariances are determined from the regression relating Y to W. To calculate this regression, the sums of squares and products for treatment effects are transferred to Table 12.3, which shows an analysis of covariance for the complete data. The only additional calculation needed to complete Table

TABLE 12.1

ORIGINAL DATA ON THE RELATIVE POTENCY OF NEW INTERNATIONAL STANDARD INSULIN IN HIMALAYAN RABBITS WHEN INJECTED ON TWO SUCCESSIVE DAYS. DATES OF INJECTION: 30/6–1/7, 9–10/7, 28–29/7 AND 6–7/8, 1936, WITH EXCEPTIONS NOTED BELOW. BODY WEIGHTS AS ON 29/7.

(From Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

Day	Treatment Log 100 × dose	Initial blood sugar level (W) in mg. per 100 mls. Data on Rabbit No. 1	2	3	4	5	6	7	8	9	10	11	12	Total
1	0·32	75	91	97	77	87	88	87	89	89	99	96	90	1,065
	0·47	94	86	99	78	101	89	81	89	85	94	93	89	1,074
	0·62	107	83	90	87	103	79	91	95	125	93	90	78	1,121
	0·77	94	93	91	93	94	81	94	84	84	89	87	86*	1,070
2	0·32	86	95	102	97	92	107	93	92	135	92	86	96*	1,173
	0·47	115	89	91	101	95	96	93	86	100	94	101	101	1,162
	0·62	100	95	94	95	102	98	94	106	85	93	101	87	1,150
	0·77	96	86	93	80	107	91	91	95	90	97	98	95	1,119
	Total	767	718	757	708	781	725	724	736	793	751	752	722	8,934

* Test on 11–12/6 instead of 6–7/8.

Mean % fall in blood sugar over 5 hours (Y)

Day	Log 100 × dose	Data on Rabbit No. 1	2	3	4	5	6	7	8	9	10	11	12	Total
1	0·32	−4·0	11·2	21·2	18·7	2·8	27·2	25·1	25·8	2·2	28·3	23·7	−2·2	180·0
	0·47	26·2	16·5	23·2	25·6	12·7	39·8	28·4	40·0	−2·4	20·4	10·3	9·2	249·9
	0·62	32·7	14·0	28·9	40·2	35·1	36·2	37·8	39·4	18·9	30·5†	30·9	6·7	351·3
	0·77	33·2	31·8	27·5	48·1	37·2	47·7	36·0	50·7	20·2†	32·4	41·8	24·2	430·8
2	0·32	19·8	21·7	26·1	32·2	28·5	35·7	20·2	26·1	5·5	22·0	14·4	20·2	272·4
	0·47	37·7	24·5	38·0	40·7	29·3	48·1	45·6	35·3	14·2	18·8	17·6	7·9	357·7
	0·62	46·8	28·6	32·6	45·4	28·6	50·4	47·7	50·0	10·4	38·1†	39·0	12·4	430·0
	0·77	44·2	31·3	44·7	48·7	38·5	52·1	39·3	63·3	12·7†	45·2	37·3	43·2	505·5
	Total	236·6	179·6	242·2	299·6	212·7	342·2	280·1	330·6	81·7	235·7	215·0	121·6	2,777·6
Body weight in kg.		2·55	2·34	1·76	1·83	1·50	1·90	1·85	1·75	2·53	2·22	1·88	1·90	24·01
Sex		F	F	F	M	F	M	M	M	F	F	M	M	

† Test on 2–3/7 instead of 30/6–1/7.

TABLE 12.2

FACTORIAL ANALYSIS OF TREATMENT COMPONENTS IN DATA FROM TABLE 12.1 WITH CORRECTION BY COVARIANCE; TREATMENT EFFECTS NUMBERED AS IN TABLE 10.3

(From Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

Source of variation	Factorial coefficients (k) for log dose: First day 0·32	0·47	0·62	0·77	Second day 0·32	0·47	0·62	0·77	Divisor $n_p S k^2$	Sums of Products of k with totals $Sk t_p$	$Sk T_p$	Variances and covariances Sw^2	Swy	Sy^2	Adjusted $Sy^2{}_a$
(1)	−1	−1	−1	−1	+1	+1	+1	+1	96	274	353·6	782·04	1,009·23	1,302·43	747·11
(2)	−3	−1	+1	+3	−3	−1	+1	+3	480	−112	1,625·4	26·13	−379·26	5,504·01	5,744·07
(3)	+3	+1	−1	−3	−3	−1	+1	+3	480	−236	−82·2	116·03	40·42	14·08	
(4)	+1	−1	−1	+1	+1	−1	−1	+1	96	−80	−0·2	66·67	0·17	0·00	
(5)	−1	+1	+1	−1	+1	−1	−1	+1	96	40	−19·4	16·67	−8·08	3·92	
(6)	−1	+3	−3	+1	−1	+3	−3	+1	480	−154	−37·2	49·41	11·94	2·88	
(7)	+1	−3	+3	−1	−1	+3	−3	+1	480	118	69·6	29·01	17·11	10·09	
Totals t_p for 12 rabbits	1,065	1,074	1,121	1,070	1,173	1,162	1,150	1,119	For adjusting Sy^2			0·098033	−0·626204	1	
T_p	180·0	249·9	351·3	430·8	272·4	357·7	430·0	505·5							

TABLE 12.3

ANALYSIS OF COVARIANCE FOR THE DATA IN TABLE 12.1 MEASURING THE RELATIVE POTENCY OF INSULIN IN RABBITS ON TWO SUCCESSIVE DAYS OF STARVATION

(Adapted from Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

Variation due to	Degrees of freedom	Sums of squares and products			Adjusted Sy^2			
		Sw^2	Swy	Sy^2	Degrees of freedom	Sum of squares	Mean square	$\sqrt{Sy_a^2}$
Rabbits	11	982·38	−1,304·29	8,383·79	11	9,296·85	845·17	
(1) Days	1	782·04	1,009·23	1,302·43	1	747·11	747·11	27·33=D
(2) Slope	1	26·13	−379·26	5,504·01	1	5,744·07	5,744·07	75·79=B
(3) Parallelism	1	116·03	40·42	14·08	1	0·14	0·14	
(4–7) Curvature	4	161·76	21·13	16·89	4	19·52	4·88	
Error	77	5,677·28	1,777·57	3,706·07	76	3,149·51	41·44	6·44=s
Total	95	7,745·62	1,164·80	8,927·27				
Coefficients for adjusting Sy^2		0·098033	−0·626204	1	$b_w=\frac{1,777·57}{5,677·28}=0·313102$			

12.3 is the variation ascribable to differences between rabbits, with which are associated 11 degrees of freedom. The method of calculating these sums of squares and products should by now be apparent—they are, of course, calculated from the total responses of each rabbit and the total initial blood sugar levels. In Table 12.3, treatment effects Nos. 4–7 are grouped together, since we are not really interested at the moment in knowing anything more about the combined regression line than whether it sufficiently describes the trend of the responses and whether the two separate lines composing it are substantially parallel.

The 96 separate responses in the test, representing 95 degrees of freedom, have already been employed in calculating statistics absorbing 18 degrees of freedom. We therefore have 77 degrees of freedom for the computation of error. The value of the slope to be used for correction by covariance is given by:

$$b_w = \frac{1{,}777{\cdot}57}{5{,}677{\cdot}28}$$

using the sums of squares and products from the error row only; hence $b_w = 0{\cdot}313102$. The coefficients for adjusting Sy^2, which are b_w^2, $-2b_w$ and unity, respectively, are therefore 0·098033, −0·626204 and unity.

The standard deviation of a single observation, s, is determined as usual from the error row by calculating Sy_a^2 for error and dividing by the number of degrees of freedom. The value of Sy_a^2 was 3,149·51, with which are associated 76 degrees of freedom (not 77, because we have "used up" one degree of freedom in the calculation of b_w); hence:

$$Ve = 41{\cdot}441$$

$$s = 6{\cdot}437$$

We may note that the variance of b_w, which is given by:

$$Vb_w = \frac{Ve}{Sw^2}$$

is $\frac{41{\cdot}441}{5{,}677{\cdot}28}$, whence: $\sigma_{b_w} = 0{\cdot}0854$

The observed regression of the percentage level of blood sugar on the initial level of blood sugar is thus highly significant. Were it not significant there would be no justification for correction by covariance.

Rabbits with a higher normal blood sugar tended to be less

sensitive to the effects of insulin than their fellows, and correction for mean initial differences between rabbits therefore accentuates the differences in sensitivity. The combined effect of insulin treatment and starvation significantly raises the normal level of blood sugar on the second day. Hence the apparent greater potency of the same dose of insulin on the second day might have been due to the general rise in blood sugar. However, although the adjusted Sy_a^2 for the effect of day of injection was reduced by correction for covariance to slightly over half its original value, it remained highly significant. Insulin therefore had an unquestionably greater potency on the second day, even when the change in the initial level of blood sugar was discounted. The mean squares 3–7, measuring departures from parallel rectilinear dose-response relationships, were less than the experimental error in both the corrected and the uncorrected material. Hence the assay is to be regarded as satisfactory and relative potency can be determined from the adjusted D^2 and B^2.

12.3. The estimation of relative potency

From the equations on p. 95 we estimate the log ratio of potencies as:

$$M=\frac{\sqrt{5}\times 0{\cdot}15051\times 27{\cdot}333}{75{\cdot}790}=0{\cdot}1214$$

The error of this estimate is:

$$s_M=\frac{6{\cdot}4375\times\sqrt{5}\times 0{\cdot}15051(\sqrt{5{\cdot}744{\cdot}07+747{\cdot}11})}{5{,}744{\cdot}07}=0{\cdot}0304$$

The same dose of insulin was therefore 1·3224 times as potent on the second day as on the first day. At the level $P=0{\cdot}05$, the limits within which this ratio has been determined are the antilogs of $0{\cdot}124\pm 0{\cdot}0304t$. The value of t with 76 degrees of freedom at the 0·05 level is 2·000; hence the limits of error are 1·150 and 1·522. Since:

$$b=\frac{2B}{I\sqrt{n_p Sk^2}}=\frac{2\times 75{\cdot}790}{0{\cdot}15051\sqrt{480}}=45{\cdot}968$$

and its standard error is:

$$\frac{s}{\sqrt{SSn_p\bar{x}_p{}^2}}=\pm 2{\cdot}761,$$

the value of $\frac{b}{s_b}$ far exceeds that at which it is necessary separately to compute exact fiducial limits.

12.4. Tests for consistency

Referring to Table 12.3 we see that the variation due to differences between rabbits is highly significant with an F of more than 20. Segregation of these differences in the analysis, made possible by the adequacy of the experimental design, eliminates errors due to the variations in the level of response between different rabbits. It does not, however, ensure that each rabbit has reacted similarly to the others despite differences which it may show from the average level of response. It is often worth while to test how consistently the animals used in an experiment have reacted to the range of doses to which all of them have been subjected.

We have so far always treated those sources of error which could not be segregated in the design of the experiment or were not purposely segregated as equivalent to variation due to random sampling, to be used as the term for experimental error. In statistical terminology, however, this error term often consists of *interactions* between experimental treatments and restrictions in design which remain after the sums of squares for these particular factors have been subtracted from the total sum of squares. In theory, each single degree of freedom available in the remaining sum of squares used as the error term in a test like the present example has associated with it a particular comparison of the mingled effects of treatment and restrictions in design. These mixed effects are known as interaction. In this case they measure the extent to which the type of response to insulin varied from one animal to another. Thus the rabbits may have differed in the slope of their separate dose-response lines and the elimination of particular relatively insensitive individuals with low slopes might be worth while in future tests, since in testing the potency of preparations of insulin we can use the same animal time and time again.

If we desire to test this possibility, the interaction between individual test animals and the non-significant treatment factors, such as those measuring departures from parallelism and rectilinearity, may serve as a restricted experimental error by which to assess the importance of those interactions which have been isolated. Table 12.4 shows the computation of the interaction between rabbits and each of the first two treatment effects shown in Table 12.2, both of which are significant. The factorial coefficients for effects 1 and 2 were applied separately to the responses of each animal to obtain the individual sums of products in the top two rows of Table 12.2. All the sums of products in Table 12.4 were

TABLE 12.4

SUMS OF PRODUCTS WITH FACTORIAL COEFFICIENTS IN TABLE 12.2 FOR INDIVIDUAL RABBITS USED IN COMPUTING INTERACTIONS BETWEEN RABBITS AND EACH OF FIRST TWO TREATMENT EFFECTS (COLUMNS 2–5); ORIGINAL AND ADJUSTED Sy^2 FOR TREATMENT EFFECTS 3–7 IN EACH RABBIT (COLUMNS 6–7)

(From Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

Rabbit No.	Treatment effect (1)		Treatment effect (2)		Treatment effects (3)–(7)	
					Original	Adjusted
	SkW	SkY	SkW	SkY	Sy^2	Sy^2
1	27	60·4	85	200·4	373·71	193·04
2	12	32·6	−18	92·2	96·33	59·41
3	3	40·6	−51	75·0	85·92	113·10
4	38	34·4	0	157·0	76·72	13·52
5	11	37·1	75	154·9	328·87	355·66
6	59	40·4	−73	124·4	69·77	79·25
7	18	25·5	26	101·5	322·75	320·84
8	22	18·8	20	200·4	112·45	109·84
9	27	3·9	−125	93·1	224·24	169·09
10	1	12·5	−17	111·3	255·62	216·48
11	20	1·6	6	165·0	340·17	343·18
12	36	45·8	−40	150·2	535·93	483·53
Total	274	353·6	−112	1,625·4	2,822·48	2,456·94
Sk^2	8	8	40	40		

then squared, the squares totalled and divided by Sk^2, the sum of the squares of the factorial coefficients, to determine the sums of squares and, since each total includes both the direct effect of treatment and its interaction with individual rabbits, the direct effect was subtracted from the total to obtain the value for interaction alone. The interaction between days and rabbits for initial blood sugars is:

$$Sw^2=\frac{27^2+12^2+\ \ldots\ 36^2}{8}-782{\cdot}04=363{\cdot}21$$

Then SkW was multiplied by SkY for each entry corresponding to individual rabbits and the product summed and the total interaction determined as for Sw^2. Thus,

$$Swy=\frac{27\times60{\cdot}4+12\times32{\cdot}6+\ \ldots\ +36\times45{\cdot}8}{8}-1{,}009{\cdot}23=105{\cdot}01,$$

and $$Sy^2=\frac{60{\cdot}4^2+32{\cdot}6^2+\ \ldots\ +45{\cdot}8^2}{8}-1{,}302{\cdot}43=425{\cdot}89.$$

The other interactions were isolated by analogous methods and Table 12.5 was constructed from the results. In it, the interaction between rabbits and (1) differences between days, and (2) the slope of the dose-response line are compared with an error term calculated

TABLE 12.5

COMPARISON OF THE INTERACTION BETWEEN RABBITS AND THE FIRST TWO TREATMENT EFFECTS WITH THE REMAINING INTERACTIONS IN TERMS OF THE REDUCED Sy^2

(From Bliss and Marks, *Quart. J. Pharm. Pharmacol.*, **12**, 182, 1939)

Interaction between rabbits and	Degrees of freedom	Sums of squares and products			Reduced Sy^2		
		Sw^2	Swy	Sy^2	Degrees of freedom	Sum of squares	Mean square
(1) Difference between days	11	363·21	105·01	425·89	11	395·53	35·96
(2) Slope of the dose-effect curve	11	967·12	433·89	488·67	11	315·21	28·66
(3–7) Remainder or "error"	55	4,346·95	1,238·67	2,791·51	54	2,438·55	45·16
Total interaction	77	5,677·28	1,777·57	3,706·07			
(1) Days+"error"	66	4,710·16	1,343·68	3,217·40	65	2,834·08	
(2) Slope+"error"	66	5,314·07	1,672·56	3,280·18	65	2,753·76	

by subtracting these two particular items of interaction from the total interaction, comprising the error term in Table 12.3. Since this is to be a critical test of significance, we must calculate reduced Sy^2s by the method described in Chapter 8.5. There was no need to test the significance of our estimates of D and B, since the value of F for the lesser of them was 18·0, which, with one and 77 degrees of freedom respectively, exceeds the 0·01 level of significance. Reduced comparisons need only be made if the variance ratio falls between values giving a P of 0·05 and 0·01. We see that the mean reduced Sy^2s are less than their experimental error and we may conclude, therefore, that despite their wide differences in overall susceptibility, all rabbits reacted similarly to the differences between days and dosages. If a significant interaction had been detected the discrepant individuals would have been sought in Table 12.4 from the products SkY for individuals.

Atypical individuals may not only reduce the precision of an assay by differing from the others in the slope of their dose-response lines, but also by reacting very erratically and thus enlarging the experimental error. Erratic reactors will tend to increase experimental error by departing in their reactions from the parallel rectilinear dose-response relationship on which the assay is based. When the present experiment was examined with this possibility in view, it was noted that experimental error for the dose-response line on the first day in rabbits 1–8 was 28·48, as compared with the equivalent 41·44 for the full experiment. This observation occurred because part of the test had been used separately as an example of the analysis of biological assays. Bliss and Marks therefore tested the possibilities that either one or more of the four additional rabbits reacted erratically to insulin, or that the response on the second day was less consistent than on the first day.

The first possibility was tested by computing for each rabbit the sum of squares measuring departure from a parallel rectilinear dose-response relationship. The factorial coefficients in rows 3–7 of Table 12.2 were applied separately to the eight percentage falls in blood sugar for each rabbit and the 12 series of products SkY were obtained, and from them Sy^2, with five degrees of freedom for each rabbit. Since unequal initial blood sugars caused part of the variability in response, each Sy^2 shown in the penultimate column of Table 12.4 was adjusted by covariance for the corresponding individual Sw^2 and Swy, using a regression coefficient (0·28114) computed from the Sw^2 and Swy for all 12 rabbits. The adjusted

Sy_a^2 thus determined for each animal is shown in the last column of Table 12.4. It was seen that eight of the rabbits reacted more or less consistently to insulin, Nos. 5, 7 and 11 were semi-erratic, and No. 12 was the most erratic of all. It was therefore suggested that in selecting individuals for future work these four rabbits might be replaced to advantage, but so far as the present experiment was concerned the log ratio of potencies and its error was practically the same, whether computed from part or all of the data, since the reduction in the mean square for experimental error was balanced by the lower reliability of the smaller number of observations.

The second possibility was tested by comparing the reduced sums of squares for curvature in the separate dose-response lines for the first and second days and it was found that the response was quite as consistent on the second as on the first day, the reduced mean square for interaction with the quadratic and cubic terms being 32·23 on the first day and 31·23 on the second day, both with 23 degrees of freedom. Bliss further tested the possibilities that larger rabbits (the weights of rabbits are shown in Table 12.1), since they received relatively less insulin than smaller ones, showed a correspondingly smaller reaction and that smaller initial blood sugars, shown by covariance to be productive of a more pronounced response, significantly affected the results. However, within the wide limits of the experiment and with a balanced design, comparison of potencies was as accurate as if injections had been made at a constant rate in mgm./kg. body weight, but differences in initial blood sugar, together with differences in body weight, were responsible for a significant amount of the variation between individuals. The percentage fall in blood sugar per unit increase in the initial level decreased more than five times as rapidly when comparing one rabbit with another, as it increased in comparing one test with another within single individuals.

These various refinements by which an assay may be tested for internal consistency, although they may sometimes suggest improvements in methods of conducting future tests, mean a considerable amount of statistical computation, but may be worth the effort if it is desired to increase the accuracy of tests to the highest possible level and to be fully satisfied about their internal consistency. In other cases the additional statistical labour involved and the frequent finding of doubtful or negative conclusions makes it easier to aim at an increased precision by increasing the number of observations made in an adequately designed experiment.

CHAPTER 13

FURTHER DESIGNS FOR ASSAYS

13.1. General remarks

The reader should by now be sufficiently familiar with the principles to be followed in computing the ratio of potencies when comparing an unknown with a standard substance to be able to work out the details of tests and their analysis. It seems desirable, however, to give a few more examples of the general structure which the design of the experiment may take under various laboratory conditions. The particular structure of an experiment will depend on the number of doses of the standard and unknown that we wish to administer, on the number of animals or other units of test material available for the experiment and on the time at our disposal. It will also be modified according to whether repeated observations can be made using the same test object, as in the use of rabbits for testing the potency of insulin, or whether, as when an animal is killed, the test object can be used once only. Another factor will, of course, be the limits of error within which we desire our results to fall.

13.2. Latin square designs

The Latin square is a particularly useful basis for the design of tests in which it is hoped to segregate a number of sources of variation which may affect the estimate of error and in which, as in the insulin test of the previous Chapter, interaction is unlikely to be significant. It may be used singly or in replication. If we plan an assay in which two doses of the standard and two doses of the unknown are to be administered to the test objects, we can design our experiment in units consisting of 4×4 Latin squares, so that each row and each column of the square will contain examples of all four treatments. It would be a rare test in which a sufficient accuracy were obtainable by the use of a single 4×4 Latin square, i.e. in which only four animals were used per dosage group. If we had decided that 12 animals per dose should give a sufficiently

accurate answer, we could design the test using three 4×4 Latin squares as shown for example in Table 13.1.

TABLE 13.1

ARRANGEMENT OF A TEST USING THREE 4×4 LATIN SQUARES

Square No.	Animal No.	Date of Test			
		1	2	3	4
1	1	S_2	S_1	U_1	U_2
	2	U_2	U_1	S_1	S_2
	3	S_1	S_2	U_2	U_1
	4	U_1	U_2	S_2	S_1
2	5	U_1	S_1	U_2	S_2
	6	S_1	U_1	S_2	U_2
	7	S_2	U_2	S_1	U_1
	8	U_2	S_2	U_1	S_1
3	9	U_2	S_2	U_1	S_1
	10	S_2	U_2	S_1	U_1
	11	S_1	U_1	S_2	U_2
	12	U_1	S_1	U_2	S_2

Under certain conditions, as in this assay, in which columns represent days of injection of a drug to animals and the treatments are completed in four days, we could run the analysis by combining the columns from all three Latin squares together, the differences between the four columns representing differences between days, it being assumed that each dose of the standard and unknown can be tested on a single animal. Thus, as an example, in the assay of parathyroid extract in which a single dog can be used for successive determinations of serum calcium, but only one determination can be made per day on each animal, the rows in Table 13.1 could represent 12 different dogs, the columns four different days with the doses as shown in the Table (cf. Bliss and Rose, *Amer. J. Hygiene*, **31**, No. 3, Sec. A, 79, 1940, for a practical example). Nevertheless, we should not be precluded from analysing separately the contributions of each Latin square, each of which might perhaps represent a batch of animals, and thus could compare groups of animals one against the other.

Correction by covariance for any concomitant measure which might be thought to influence the results of the test will follow the usual scheme and need not be further elaborated. It must be

noted that the use of the Latin square precludes a full analysis of interaction, as confounding occurs; but it is fortunate that in the general run of biological tests interaction is frequently negligible.

13.3. Symmetrical pairs

Another design which may be extended to any number of treatments and which has the advantage that each test object need be used twice only is that known as symmetrical pairs. The arrangement is illustrated in Table 13.2. In this table it is supposed that 12 animals are tested on four different doses, two of the standard and two of the unknown.

TABLE 13.2

ARRANGEMENT OF ONE REPLICATION OF A TEST USING SYMMETRICAL PAIRS

Animal No.	1st Test	2nd Test	Animal No.	1st Test	2nd Test
1	U_1	U_2	7	U_2	U_1
2	U_1	S_1	8	S_1	U_1
3	U_1	S_2	9	S_2	U_1
4	U_2	S_1	10	S_1	U_2
5	U_2	S_2	11	S_2	U_2
6	S_1	S_2	12	S_2	S_1

When only two observations have been made on each test object the segregation of variation between test animals is computed in terms of the differences between the two readings on each individual. There are then the same number of differences and degrees of freedom as there are individuals and the usual correction for the mean is omitted. The effects of treatments are determined from the sums of the differences for the six combinations of treatments, with the aid of factorial coefficients. These are not the same as the factorial coefficients in other designs, and the coefficients for two and three doses on the standard and unknown are shown in Table 13.3. Each dose difference has a factorial coefficient relevant to that particular comparison and the divisor, sum of products, etc., are computed by the usual formulae. The scheme for computing the totals at the bottom of each column for factorial coefficients in Table 13.3 is exemplified in Table 13.4. Note that the divisors in Table 13.3 are formed from the quantities $n_p Sk^2$, where n_p is the total number of observations in each comparison such as $U_1 - U_2$.

TABLE 13.3

FACTORIAL COEFFICIENTS (k) FOR THE SEGREGATION OF INDIVIDUAL TREATMENT EFFECTS IN THE ANALYSIS OF ASSAYS BASED ON SYMMETRICAL PAIRS WITH 2 OR 3 DOSES OF EACH SUBSTANCE

Source of variation	U_1–U_2	U_1–S_1	U_1–S_2	U_2–S_1	U_2–S_2	S_1–S_2	Divisor n_pSk^2
(1) Differences between samples	0	+1	+1	+1	+1	0	$4n_p$
(2) Linear regression	−1	0	+1	+1	0	−1	$4n_p$
(3) Departure from parallelism	−1	−1	0	0	+1	+1	$4n_p$

Source of variation	U_1–U_2	U_1–U_3	U_1–S_1	U_1–S_2	U_1–S_3	U_2–U_3	U_2–S_1	U_2–S_2	U_2–S_3	U_3–S_1
(1) Differences between samples	0	0	+1	+1	+1	0	+1	+1	+1	+1
(2) Linear regression	−1	−2	0	+1	−2	−1	+1	0	−1	+2
(3) Departure from parallelism	−1	−2	−2	+1	0	−1	−1	0	+1	0
(4) Combined curvature	+1	0	0	+1	0	−1	−1	0	−1	0
(5) Opposed curvature	+3	0	+2	−1	+2	−3	−1	−4	−1	+2

	U_3–S_2	U_3–S_3	S_1–S_2	S_1–S_3	S_2–S_3	Divisor n_pSk^2
(1)	+1	+1	0	0	0	$9n_p$
(2)	+1	0	−1	−2	−1	$24n_p$
(3)	+1	+2	+1	+2	+1	$24n_p$
(4)	+1	0	+1	0	−1	$8n_p$
(5)	−1	+2	−3	0	+3	$72n_p$

TABLE 13.4

SCHEME FOR TABULATING DIFFERENCES IN THE ANALYSIS OF AN ASSAY USING SYMMETRICAL PAIRS AND k REPLICATIONS

Pair No.	Treatment differences	Total	Pair No.	Treatment differences	Total	Totals
1	U_1-U_2 (k replications)	T_1	7	U_2-U_1 (k replications)	T_7	U_1-U_2 $(=T_1-T_7)$
2	U_1-S_1 ,,	T_2	8	S_1-U_1 ,,	T_8	U_1-S_1 $(=T_2-T_8)$
3	U_1-S_2 ,,	T_3	9	S_2-U_1 ,,	T_9	U_1-S_2 $(=T_3-T_9)$
4	U_2-S_1 ,,	T_4	10	S_1-U_2 ,,	T_{10}	U_2-S_1 $(=T_4-T_{10})$
5	U_2-S_2 ,,	T_5	11	S_2-U_2 ,,	T_{11}	U_2-S_2 $(=T_5-T_{11})$
6	S_1-S_2 ,,	T_6	12	S_2-S_1 ,,	T_{12}	S_1-S_2 $(=T_6-T_{12})$

Totals for all 12 pairs: T_a, T_b . . . T_k

If there are several replications of the experiment using, for instance, four sets each of twelve individuals, the total differences for all sets together are computed for the purposes of analysis, but we can segregate differences between sets in the analysis of variance or covariance from the totals of differences within each set. The final analysis of variance then takes the form shown in Table 13.5, with an error term based on the residual sum of squares after eliminating the differences between sets and the differences between the various treatment effects. In this Table D and B, computed from the mean square for differences between samples and slope respectively, have the same meaning as usual and the computation of relative potency and its error is made by the usual method.

TABLE 13.5

THE ANALYSIS OF VARIANCE IN A 2×2 DOSE ASSAY USING SYMMETRICAL PAIRS

Source of variation	Degrees of freedom	Sum of squares	Mean square
k replications	k	$\frac{(T_a^2+T_b^2+\ldots+T_k^2)}{24}$	$\frac{T_k^2}{24k}$
(1) Differences between samples	1	From factorial analysis	D^2
(2) Linear regression	1		B^2
(3) Departure from parallelism	1		—
Random sampling (n pairs of observations)	$n-k-3$	Computed by subtraction	$\frac{\text{Sum of squares}}{n-k-3}$
Total within pairs	n	$\frac{S(\text{diffs.})^2}{2}$	$\frac{S(\text{diffs.})^2}{2n}$

13.4. Balanced incomplete blocks

Sometimes the limitations of the experimental material, as when animals are segregable into litters each with only a small number of members, make it impossible for each treatment to be given to each

test object, but it may be possible to give more treatments per test object or set than in such a design as symmetrical pairs. Advantage of this possibility may be taken by the use of balanced incomplete block designs, in which the number of units per block (block implying a statistical unit, such as the Latin square) is less than the number of different treatments, i.e. doses, to be administered. Differences between blocks can be eliminated if the design is such that every possible combination of all treatments occurs in the same number of blocks. The handling of this design is not simple and it would be inadvisable to attempt to base tests on such a design without professional assistance. We note here the existence of the method so that if the occasion appears to warrant it, the research worker is aware of the possible application of the method to his material.

There is only a limited number of solutions of the problem presented by this technique which employ a practicable number of blocks and which are available to the assayist. These are given by Fisher and Yates (*Statistical Tables for Biological, Agricultural and Medical Research*, Oliver & Boyd, Edinburgh) for 10 replications or less. As with Latin squares, we can cope with almost any desired type of assay by one or other of these solutions, alone or in replication.

As an example, four dose levels of the standard and four of an unknown can be tested in fourteen blocks of four test objects each (e.g. four litter-mates) with seven test objects per dose. The arrangement is shown in Table 13.6.

TABLE 13.6

AN ARRANGEMENT FOR TESTING FOUR DOSES EACH OF A STANDARD AND UNKNOWN, USING BALANCED INCOMPLETE BLOCKS

Block Number	Doses of Standard and Unknown			
1	S_1	S_2	S_3	S_4
2	U_1	U_2	U_3	U_4
3	S_1	S_2	U_3	U_4
4	S_3	S_4	U_1	U_2
5	S_1	S_3	U_2	U_4
6	S_2	S_4	U_1	U_3
7	S_1	S_4	U_2	U_3
8	S_2	S_3	U_1	U_4
9	S_1	S_2	U_1	U_2
10	S_3	S_4	U_3	U_4
11	S_1	S_3	U_1	U_3
12	S_2	S_4	U_2	U_4
13	S_1	S_4	U_1	U_4
14	S_2	S_3	U_2	U_3

In general, if

b = the number of blocks,

k = the number of test objects per block,

n_p = the number of test objects per dose,

r = the number of doses of the standard and unknown together,

then $rn_p = bk$,

and $\frac{n_p(k-1)}{r-1} = \begin{cases} \text{the number of blocks shared in common by any} \\ \text{and every two doses levels.} \end{cases}$

We calculate the set of r quantities $kQ_p = kSY_p - ST_p$, where ST_p is the sum of the total responses in all blocks containing a test object with response Y_p (responding to dose X_p).

The adjusted mean responses, $\bar{Y}_p$, are given by the quantities

$$\bar{Y}_p = \bar{Y} + \frac{kQ_p(r-1)}{rn_p(k-1)},$$

where $\bar{Y}$ is the general mean.

The general scheme for the analysis of variance is discussed in Fisher & Yates (ref. above).

13.5. Confounding

We have met previously the concept of confounding two or more possible sources of variation. It is possible purposely to confound certain sources and to eliminate any effect they may have on our estimate of error as a whole. Sometimes confounding is unavoidable. When in experimentation with animals we can place only one animal in each cage or pen, the differences between animals are unavoidably confounded with differences between cages. On the other hand, when it is possible to cage groups of animals together it is not theoretically necessary for confounding to occur. In practice, it may be very much more convenient to confound differences between treatments with differences between cages, simply because it is easier to inject the same dose into all the animals in a cage than it is to pick a particular animal in each cage and give it its allotted separate dose.

There are sometimes conditions in which it is undesirable to give different doses to animals in the same cage, as in experiments where a preparation of the active material passes into the excreta and may be reingested by the animals. However, if it is at all possible to avoid this type of confounding, especially in the early stages of work and until it may be shown that differences between cages

are of no importance, the attempt should be made. In the last resort we should at least split the total dose group among two or more separate cages. There has been in biological work a considerable tendency to ignore the possibility of differences in reaction due to animals being caged in distinct groups and it seems to have been tacitly assumed that variation between cages must be negligible. It must be a rarely designed animal house in which conditions are so uniform that this assumption can be justified, and in the light of our knowledge that a variety of responses are influenced by health, temperature, light, feeding and many other factors, it would always seem worth while so to arrange our preliminary trials that the contributions of these factors to differences in the location of test objects may be examined.

When dealing with tests in which a series of relatively rapid observations of the effects of a drug takes in all a considerable time, it is also important to avoid confounding response with order of treatment. If it takes only half an hour to inject 100 animals and the total test period is several days, the order of injection need not worry us, but if each complete measurement of response follows treatment half an hour later and we must spend the whole day in getting through a complete test, it is advisable, and sometimes essential, to randomise the order in which treatments are administered or, alternatively, to arrange the test as a series of replications, just as is done when a complete test covers several days of separate treatments.

It is frequently easy to enumerate more possible sources of variation which may be of importance than can be coped with by the restrictions in design. When this is so, we must arrange for the individual segregation of those sources of variation which are known or believed to be of the greatest importance and for the confounding of those believed to be of the least importance. If the magnitude of the variation attributable to the confounded factors warrants it, they can be examined separately at a later period.

13.6. The twin cross-over test

A test designed by Smith, Marks, Fieller & Broom (*Quart. J. Pharm. Pharmacol.* **17,** 108, 1944) for the routine assay of insulin with rabbits is a neat illustration of the isolation of the most important features of a test, and the confounding of relatively unimportant sources of variation with differences between rabbits. In this test, the general design of which is of wide application, a

more accurate estimate of the difference in response to the standard and unknown and of their combined slope is obtained than of the extent of departures from parallelism of the two separate dose-response lines. This is a particularly useful design for routine work, where we are in any case reasonably certain that the slopes of the two preparations are identical.

The test is complete in two days of laboratory work, and involves four groups of animals receiving the following treatments:

Group	1st day	2nd day
1	S_2	U_1
2	S_1	U_2
3	U_2	S_1
4	U_1	S_2

where S_1 and S_2 represent respectively the low and high doses of the standard preparation and U_1 and U_2 those of the test preparation, the log dose interval being the same for each substance. There should not be less than three animals per group.

Comparisons which can be built up from differences between the reactions of individual groups on the two days of testing are made with an accuracy dependent on the variation *within rabbits*, while those which depend on differences between the reactions of different groups are made with an accuracy dependent on the variation *between rabbits*, which is normally greater than that within rabbits. Although the analysis of variance is simple when all groups contain equal numbers of rabbits, it is more informative to analyse in terms of sums and differences, since the case of unequal group numbers is then easily managed.

We write down the mean responses to each treatment as in Table 13.7 and also their sums (Y_1 to Y_4) and differences (y_1 to y_4) within groups. The variances *within groups* for the quantities Y

TABLE 13.7

ARRANGEMENT OF THE RESULTS OF A TWIN CROSS-OVER TEST FOR ANALYSIS

Group	No. of animals	Mean response to:		Sum	Difference
1	n_1	S_2	U_1	Y_1	y_1
2	n_2	S_1	U_2	Y_2	y_2
3	n_3	S_1	U_2	Y_3	y_3
4	n_4	S_2	U_1	Y_4	y_4

and y are estimated from the individual sums and differences for each separate animal, each associated with $(Sn_p - 4)$ degrees of freedom, where n_p is the number of animals in group p. These are

denoted by VY and Vy. The following quantities are also calculated:

$$\frac{1}{w}=S\frac{1}{n_p};\quad \frac{1}{w'}=-\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}-\frac{1}{n_4}$$

Departure from parallelism of the two dose-response lines for substances is indicated if the quantity $(Y_1-Y_2-Y_3+Y_4)$ significantly exceeds zero; its variance is $\frac{VY}{w}$.

The log ratio of potency is given by:

$$M=I\,Sy/-y_1+y_2+y_3-y_4,$$

where I is the log dose interval.

The fiducial limits of error are obtained by the following steps:

(i) F for 1 and (Sn_p-4) d.f. is taken from the first column of the table of F at the required probability level.

(ii) $U^2=(-y_1+y_2+y_3-y_4)^2-\frac{FVy}{w}$, which must be positive if real limits are calculable.

(iii) $UT=(-y_1+y_2+y_3-y_4)\,Sy-\frac{FVy}{w'}$

(iv) The limits are the roots of the equation:
$U^2m^2-2UTIm+T^2I^2=0$, solving for m.

A difference in slope between days 1 and 2 does not invalidate the assay.

It is to be noted that, although some of the degrees of freedom associated with differences between rabbits are confounded with differences between groups, there is nevertheless available in this test the remaining degrees of freedom from which to estimate the variance of between-group comparisons.

The test can be extended quite simply to a "triplet" design, with three dosage groups per substance. This allows departure from linearity to be detected with a variance based on the mean square between rabbits, but the analysis is relatively simple only if there are equal numbers of animals in all groups. The intermediate doses are crossed-over in the additional groups:

Group	1st day	2nd day
1	S_3	U_1
2	S_2	U_2
3	S_1	U_3
4	U_3	S_1
5	U_2	S_2
6	U_1	S_3

CHAPTER 14

DISCONTINUOUS VARIATION

14.1. Enumeration data

So far we have been dealing with the analysis of a series of responses, each of which is a separate numerical measure forming part of a theoretically continuous distribution. We now pass to considering the statistical methods applicable when the response is *quantal*. A quantal response occurs when all that we record about a group of test objects is whether or not each of its members exhibits some characteristic effect. When a poison is administered to a group of animals, some may die, while the rest remain alive and the response of that group will then be simply the proportion dying—a matter of *enumeration*. Other characteristic examples of quantal responses used in biological assay involve the presence or absence of certain changes in the blood, in the vaginal smear or in the ovaries, and the presence or absence of organisms in samples.

14.2. The Chi-squared test

If we have a number of groups of observations in which the presence or absence of a specific attribute is enumerated, we may wish to know whether or not they form a homogeneous series. Table 14.1 shows the proportions of dead and living animals in

TABLE 14.1

CALCULATION OF χ^2 IN A TEST OF TOXICITY

Group	Observed:			Expected		Differences		χ^2	
	Dead	Alive	Total	Dead	Alive	Dead	Alive	Dead	Alive
1	17	33	50	17·31	32·69	0·31	−0·31	0·056	0·029
2	14	22	36	12·46	23·54	−1·54	1·54	0·190	0·107
3	9	19	28	9·69	18·31	0·69	−0·69	0·049	0·026
4	13	21	34	11·77	22·23	−1·23	1·23	0·128	0·068
5	10	24	34	11·77	22·23	1·77	−1·77	0·258	0·137
Total	63	119	182	63·00	119·00	0·00	0·00	0·681	0·367
Percentage	34·62	65·38	100·0						

$\chi^2 = 1{\cdot}048$, degrees of freedom $= 4$, $P > 0{\cdot}50$

TABLE 14.2

DISTRIBUTION OF χ^2

	Probability												
n	·99	·98	·95	·90	·80	·70	·50	·30	·20	·10	·05	·02	·01	·001
1	$0{\cdot}0^3157$	$0{\cdot}0^3628$	0·00393	0·0158	0·0642	0·148	0·455	1·074	1·642	2·706	3·841	5·412	6·635	10·827
2	0·0201	0·0404	0·103	0·211	0·446	0·713	1·386	2·408	3·219	4·605	5·991	7·824	9·210	13·815
3	0·115	0·185	0·352	0·584	1·005	1·424	2·366	3·665	4·642	6·251	7·815	9·837	11·345	16·268
4	0·297	0·429	0·711	1·064	1·649	2·195	3·357	4·878	5·989	7·779	9·488	11·668	13·277	18·465
5	0·554	0·752	1·145	1·610	2·343	3·000	4·351	6·064	7·289	9·236	11·070	13·388	15·086	20·517
6	0·872	1·134	1·635	2·204	3·070	3·828	5·348	7·231	8·558	10·645	12·592	15·033	16·812	22·457
7	1·239	1·564	2·167	2·833	3·822	4·671	6·346	8·383	9·803	12·017	14·067	16·622	18·475	24·322
8	1·646	2·032	2·733	3·490	4·594	5·527	7·344	9·524	11·030	13·362	15·507	18·168	20·090	26·125
9	2·088	2·532	3·325	4·168	5·380	6·393	8·343	10·656	12·242	14·684	16·919	19·679	21·666	27·877
10	2·558	3·059	3·940	4·865	6·179	7·267	9·342	11·781	13·442	15·987	18·307	21·161	23·209	29·588
11	3·053	3·609	4·575	5·578	6·989	8·148	10·341	12·899	14·631	17·275	19·675	22·618	24·725	31·264
12	3·571	4·178	5·226	6·304	7·807	9·034	11·340	14·011	15·812	18·549	21·026	24·054	26·217	32·909
13	4·107	4·765	5·892	7·042	8·634	9·926	12·340	15·119	16·985	19·812	22·362	25·472	27·688	34·528
14	4·660	5·368	6·571	7·790	9·467	10·821	13·339	16·222	18·151	21·064	23·685	26·873	29·141	36·123
15	5·229	5·985	7·261	8·547	10·307	11·721	14·339	17·322	19·311	22·307	24·996	28·259	30·578	37·697

16	5·812	6·614	7·962	9·312	11·152	12·624	15·338	18·418	20·465	23·542	26·296	29·633	32·000	39·252
17	6·408	7·255	8·672	10·085	12·002	13·531	16·338	19·511	21·615	24·769	27·587	30·995	33·409	40·790
18	7·015	7·906	9·390	10·865	12·857	14·440	17·338	20·601	22·760	25·989	28·869	32·346	34·805	42·312
19	7·633	8·567	10·117	11·651	13·716	15·352	18·338	21·689	23·900	27·204	30·144	33·687	36·191	43·820
20	8·260	9·237	10·851	12·443	14·578	16·266	19·337	22·775	25·038	28·412	31·410	35·020	37·566	45·315
21	8·897	9·915	11·591	13·240	15·445	17·182	20·337	23·858	26·171	29·615	32·671	36·343	38·932	46·797
22	9·542	10·600	12·338	14·041	16·314	18·101	21·337	24·939	27·301	30·813	33·924	37·659	40·289	48·268
23	10·196	11·293	13·091	14·848	17·187	19·021	22·337	26·018	28·429	32·007	35·172	38·968	41·638	49·728
24	10·856	11·992	13·848	15·659	18·062	19·943	23·337	27·096	29·553	33·196	36·415	40·270	42·980	51·179
25	11·524	12·697	14·611	16·473	18·940	20·867	24·337	28·172	30·675	34·382	37·652	41·566	44·314	52·620
26	12·198	13·409	15·379	17·292	19·820	21·792	25·336	29·246	31·795	35·563	38·885	42·856	45·642	54·052
27	12·879	14·125	16·151	18·114	20·703	22·719	26·336	30·319	32·912	36·741	40·113	44·140	46·963	55·476
28	13·565	14·847	16·928	18·939	21·588	23·647	27·336	31·391	34·027	37·916	41·337	45·419	48·278	56·893
29	14·256	15·574	17·708	19·768	22·475	24·577	28·336	32·461	35·139	39·087	42·557	46·693	49·588	58·302
30	14·953	16·306	18·493	20·599	23·364	25·508	29·336	33·530	36·250	40·256	43·773	47·962	50·892	59·703

For larger values of n, the expression $\sqrt{2\chi^2}-\sqrt{2n-1}$ may be used as a normal deviate with unit variance, remembering that the probability for χ^2 corresponds with that of a single tail of the normal curve.

Table 14.2 is reprinted from Table IV of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

five different groups of rats, each receiving a standard dose of nembutal. Nembutal is an anaesthetic which, if given in too large an amount, causes death. We see that the percentage of dead rats in each group varies between 29·1 % and 38·9 %, and therefore wish to determine whether these five groups can be regarded as five samples of the same population.

We approach the problem on the null hypothesis. The best estimate we have of the death rate in the homogeneous population from which we will suppose the groups to have been drawn is the mean death rate calculated from the totals of all groups. These are 63 dead and 119 living, or 34·62 % dead. We then use this percentage to predict in each group the expected number of dead and we write this expected number, together with the corresponding expected number living, in Table 14.1. We now proceed to calculate the statistic χ^2, where

$$\chi^2=\frac{(Y-E)^2}{E}$$

Y is the observed number living or dying and E the expected number living or dying, and χ^2 must be calculated for both classes. Since we calculate the value of χ^2 for each class separately, χ^2 for each group is written in the last two columns of Table 14.1. The ten values of χ^2 are then summed, since χ^2 is additive, and from a Table of χ^2 (Table 14.2) entered with the appropriate number of degrees of freedom, we test whether or not χ^2 exceeds in value a reasonable figure.

The number of degrees of freedom in a Table of this type, when testing a null hypothesis, are, as usual, the number of independent comparisons which may be made. Thus, although we have added together ten separate determinations of χ^2 to determine the final value, 1·048, there are only four degrees of freedom, since there are only four independent comparisons possible between the proportions dying in the different groups. The proportions of living are merely the proportions of dying subtracted from 1 (the percentages living being the percentages dying subtracted from 100). We therefore enter the Table of χ^2 with four degrees of freedom and see that a figure as high as 3·36 may be expected in 50 % of cases, and hence that no significance can be attached to the value found. We conclude that the five groups are five samples from the same homogeneous population.

A shortened form of this calculation applicable to the type of

data under consideration is given in Chapter 15.3. The extended form used here exemplifies the general theory, and the shortened form is based on the identity:

$$\frac{(Y-E)^2}{E}+\frac{(Y-E)^2}{n_p-E}=\frac{n_p(Y-E)^2}{E(n_p-E)}$$

When dealing with graded responses we would have calculated the variance between and within groups and compared the two values by an F-test to arrive at this type of conclusion. χ^2 is in fact n_1F, where n_1 is the number of degrees of freedom associated with the greater mean square in an F-test in which n_2 is infinite. χ^2 should not be calculated from very small classes. If the expectation in any cell of a Table is less than 5, we should combine the data in that cell with the data in another cell, if possible, to obtain an expectation greater than 5.

14.3. The binomial distribution

If p is the proportion of members of a population which show a characteristic effect and $q=1-p$ is the proportion not showing the effect, then the number of members in a sample of n members showing the effect in question will be 0, 1, 2, 3, etc., in the respective proportions of cases q^n, $nq^{n-1}p$, $\frac{n(n-1)}{2}q^{n-2}p^2$ . . . where in general r individuals will show the effect p is a proportion of samples given by the formula:

$$\frac{q^{n-r}p^r\,|\,\underline{n}}{\underline{r}\,|\,\underline{n-r}}$$

Readers familiar with the binomial theorem will recognise that these are the successive terms of the binomial expansion of $(q+p)^n$. The mean of such a series, when measuring the number showing the characteristic effect, is np and the variance exhibited by samples of n is npq. The variance is thus correlated with the mean and is greatest when $p=q=0{\cdot}5$. We could test the homogeneity of a series of results, such as those in Table 14.1, by utilising this theorem, but the χ^2 test is more adequate and easier to apply. Once more we see how the parameters of populations from which groups exhibiting quantal responses are drawn are calculable on theoretical grounds alone. We have no practical measure of the variance within groups, but we know what this variance should be from the binomial theorem.

14.4. The dose-response relationship

If a series of doses of a preparation is given to groups of test objects and the response of each group is the percentage showing a characteristic effect, we may relate dose to response by plotting percentage responses against dose. If we do this we shall obtain an *S*-shaped curve, illustrated in Figures 14.1 and 14.2. This

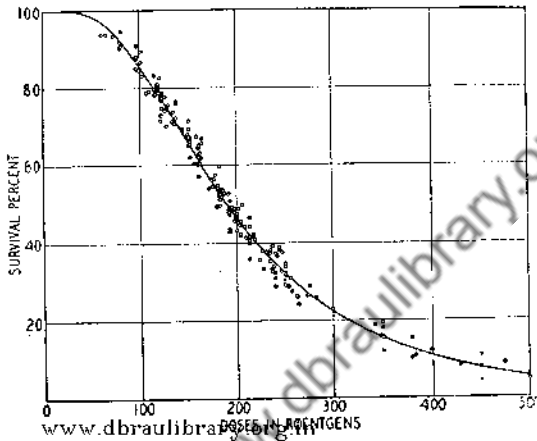


FIG. 14.1. The percentage of *Drosophila* eggs surviving various doses of roëntgen rays. (From Bliss and Packard, *Amer. J. Roentg. & Rad. Therap.*, **46**, 400, 1941.)

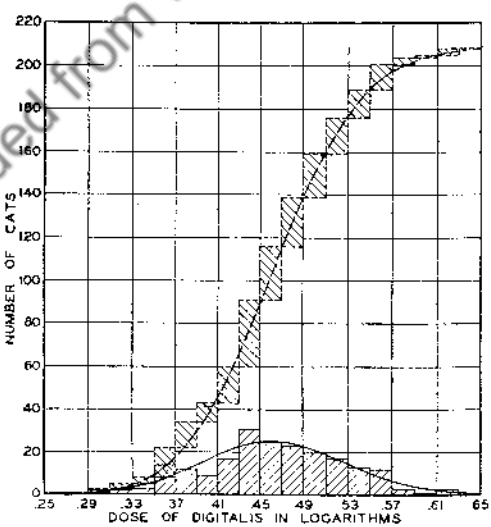


FIG. 14.2. Initial and cumulative frequency distributions of individual lethal doses in logarithms of a preparation of digitalis administered to cats. (From Bliss and Hanson, *J. Amer. Pharm. Assoc.*, **28**, 521, 1939.)

curve is still *S*-shaped if we plot the response against the logarithm of the dose, although in the majority of cases the upper and lower arms of the *S* then exhibit similar curvature and the curve is thus symmetrical about the mid-point. A special transformation is necessary which, as long as a similar relationship holds between log dose and response as often does with graded responses, will in theory convert this curve into a straight line.

To understand the implication of this transformation we must consider the effect of individual responses. The smallest dose which

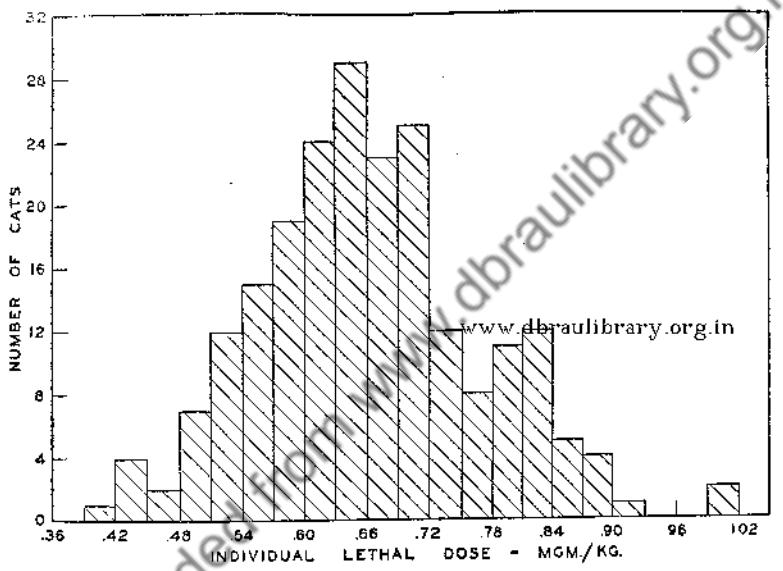


FIG. 14.3. Frequency distribution of the individual lethal doses of a tincture of digitalis administered to cats. (From Bliss, *J. Amer. Pharm. Assoc.*, **33**, 225, 1944.)

will produce the characteristic effect in any given animal is the *individual effective dose*. In a population of test objects the individual effective dose will vary from object to object. The distribution of individual effective doses will be represented by a frequency curve in which the abscissa is the log dose and the area of the curve to the left of the ordinate represents the proportion of animals which will respond to any particular log dose (Figures 14.3 and 14.4). The transformation we are about to discuss assumes that this distribution will be a normal distribution, or, in so many words, that the logarithms of the individual effective

doses are normally distributed. This theoretical expectation is often fulfilled in practice. If we call the logarithm of the individual

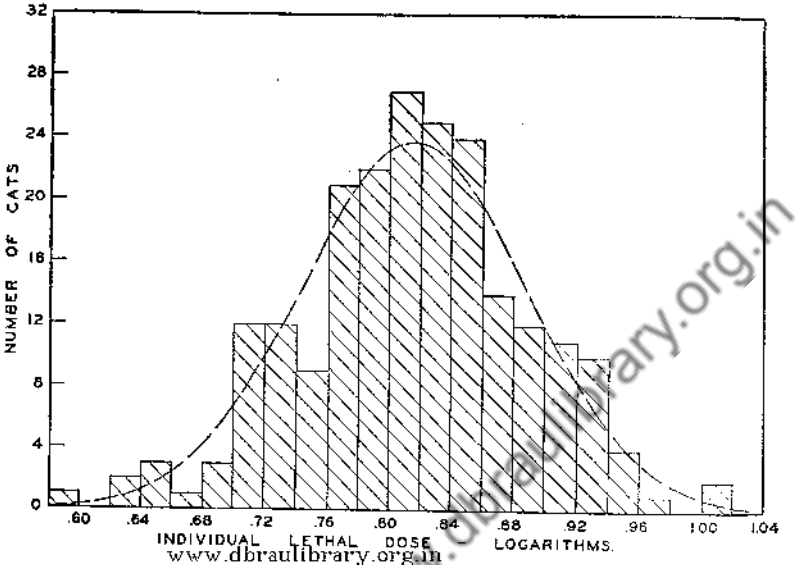


FIG. 14.4. The frequency distribution in Fig. 14.3 plotted on a logarithmic dose scale. The smooth curve is that expected from a theoretical normal distribution. (From Bliss, *J. Amer. Pharm. Assoc.*, 33, 225, 1944.)

effective dose X, $\bar{X}$ is its mean value in a large population of test objects and the quantity

$$Y = \frac{X - \bar{X}}{\sigma_x}$$

is called the normal equivalent deviation. Y is related to P, the proportion of animals reacting when given the dose X, by the equation:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{1}{2}Y^2} dY$$

If we know the value of P we can calculate Y, and by determining the values of X and σ we can estimate the linear dose-response relation between X and Y and the errors relevant to it. The goodness of fit of our original observations to the calculated line may be judged by a modified form of χ^2 test.

Tables of the normal equivalent deviation corresponding to percentages from 0–100 have been calculated, but since for per-

centages below 50 it has a negative value, it is more convenient to add a constant to Y large enough to make all the practical values positive. The constant adopted was 5 and the quantity $(5+Y)$ is called the *probit* and is used in calculations just as would be the normal equivalent deviation itself. A normal equivalent deviation of -5 equals a probit of 0 and corresponds to an extremely small percentage of test objects showing the characteristic effect. Figures 14.5 and 14.6 show the dose-response lines in Figures 14.1 and 14.2 plotted as probits against log dose.

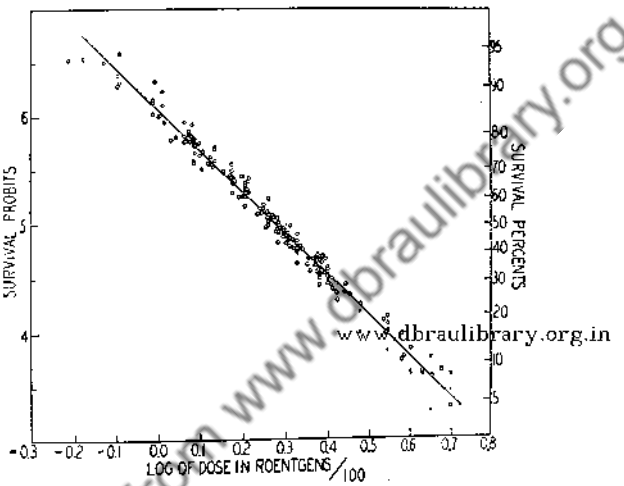


FIG. 14.5. The dosage-survival curve of Fig. 14.1 plotted as probit against log dose. (From Bliss and Packard, *Amer. J. Roentg. & Rad. Therap.*, **46**, 400, 1941.)

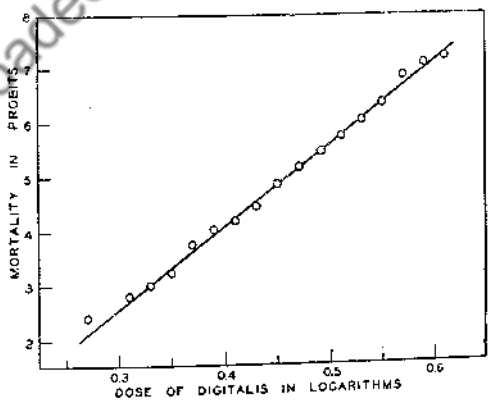


FIG. 14.6. The cumulative curve of Fig. 14.2 plotted as probit against log dose. (From Bliss and Hanson, *J. Amer. Pharm. Assoc.*, **28**, 521, 1939.)

14.5. The variance of a probit

We have seen that the variance of a binomial distribution is correlated with the mean and is smallest when p approaches 0 or 1. In transforming proportions or percentages to probits we "stretch out" the tails of the distribution to such an extent that the variance of a probit is least when $p=0{\cdot}5$ and increases as p approaches 0 or 1, i.e. as normal equivalent deviation approaches $-\infty$ and $+\infty$. In terms of percentages, if P and Q are the percentages of reacting and non-reacting objects in a group,

$$VP=\frac{PQ}{n_p} \text{ and } VY=\frac{PQ}{n_pZ^2}, \text{ where } Z=\frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}Y^2}$$

VY is the variance of the probit. The log dose-response line is fitted by least squares, but since the variance of a probit is not constant for all values it may assume, different groups do not have equal weight even when the same number of observations are made per group. The weight of an observation is inversely proportional to the variance and is thus:

$$\frac{n_pZ^2}{PQ}=n_pw$$

where n_p is the number of animals in the group and w is the *weight factor*. The straight line is then fitted by finding the minimal value of $Sn_pw(Y-E)^2$, where E is the estimated value of Y, as in fitting graded responses by the principle of least squares.

14.6. The estimation of the dose-response line

The dose-response line is estimated by the use of successive approximations. To start with, a *provisional line* is fitted either by eye or by approximate computation. If the reader decides that he cannot fit a good enough provisional line graphically, he may use weights and probits corresponding to the actual observations in the calculation of a provisional line. It is unfortunate that the method of fitting the dose-response line involves a series of approximations, and a little trouble over obtaining a good initial fit is worth while, as with a good provisional line it is often unnecessary to repeat the procedure more than once. Naturally, if the first calculated line practically coincides with the provisional line there is no need for further computation.

Probits corresponding to the crude observations are called *empirical probits* and from the provisional line we obtain, by reading

them off from each dose level, the values of the *expected probits*. From the empirical and expected probits we compute at each dose level a *corrected probit*, Y_c, which is used in the actual calculation of the regression line. This procedure may be used not only with percentage responses lying between 0 and 100 but also when no reactions are observed or when all the animals give the characteristic response. The empirical probits for 0 or 100% response are $-\infty$ and $+\infty$, but the expected probits obtained from the provisional line, using the remaining points with, if possible, some allowance in the graphical fitting for the presence of 0 or 100% responses at other dosage levels, will be finite values.

The equation for computing a corrected probit is:

$$Y_c = Y + \frac{Q}{Z} - \frac{q}{Z}$$

where Y_c is the corrected probit, Y is the expected probit as read from the provisional line, Q is the area of the tail of the normal curve beyond the point Y, Z is the ordinate of the normal curve at the point Y, and q is the observed proportion of non-reacting test objects. When q is 0, the corrected probit is $Y + \frac{Q}{Z}$, and this value of Y_c is called the *maximum corrected probit*. The inverse relationship:

$$Y_c = Y - \frac{P}{Z} + \frac{p}{Z}$$

gives the same results as above, using p, the proportion of reactors in a group, and when p is 0 the value of

$$Y_c = Y - \frac{P}{Z}$$

gives us the *minimum corrected probit*. The quantity $\frac{1}{Z}$ by which we correct the empirical probit is known as the *range*. Table 14.3 gives the values of probits for transforming the dosage mortality curve to a straight line for various values of the percentage reactors. It is normally accurate enough, particularly when the percentage reacting falls between 10 and 90, to take the percentage to two or at the most three significant figures. Table 14.4 lists maximum and minimum corrected probits, the range and the weighting coefficients for each value of the expected probit in steps of 0·1.

TABLE 14.3

PROBITS

Transformation of the Sigmoid Dosage Mortality Curve to a Straight Line. (C. I. Bliss)

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1	2	3	4	5
0	—	1·9098	2·1218	2·2522	2·3479	2·4242	2·4879	2·5427	2·5911	2·6344					
1	2·6737	2·7096	2·7429	2·7738	2·8027	2·8299	2·8556	2·8799	2·9031	2·9251					
2	2·9463	2·9665	2·9859	3·0046	3·0226	3·0400	3·0569	3·0732	3·0890	3·1043	For more detail see				
3	3·1192	3·1337	3·1478	3·1616	3·1750	3·1881	3·2009	3·2134	3·2256	3·2376	values for 95–100.				
4	3·2493	3·2608	3·2721	3·2831	3·2940	3·3046	3·3151	3·3253	3·3354	3·3454					
5	3·3551	3·3648	3·3742	3·3836	3·3928	3·4018	3·4107	3·4195	3·4282	3·4368	9	18	27	36	45
6	3·4452	3·4536	3·4618	3·4699	3·4780	3·4859	3·4937	3·5015	3·5091	3·5167	8	16	24	32	40
7	3·5242	3·5316	3·5389	3·5462	3·5534	3·5605	3·5675	3·5745	3·5813	3·5882	7	14	21	28	36
8	3·5949	3·6016	3·6083	3·6148	3·6213	3·6278	3·6342	3·6405	3·6468	3·6531	6	13	19	26	32
9	3·6592	3·6654	3·6715	3·6775	3·6835	3·6894	3·6953	3·7012	3·7070	3·7127	6	12	18	24	30
10	3·7184	3·7241	3·7298	3·7354	3·7409	3·7464	3·7519	3·7574	3·7628	3·7681	6	11	17	22	28
11	3·7735	3·7788	3·7840	3·7893	3·7945	3·7996	3·8048	3·8099	3·8150	3·8200	5	10	16	21	26
12	3·8250	3·8300	3·8350	3·8399	3·8448	3·8497	3·8545	3·8593	3·8641	3·8689	5	10	15	20	24
13	3·8736	3·8783	3·8830	3·8877	3·8923	3·8969	3·9015	3·9061	3·9107	3·9152	5	9	14	18	23
14	3·9197	3·9242	3·9286	3·9331	3·9375	3·9419	3·9463	3·9506	3·9550	3·9593	4	9	13	18	22
15	3·9636	3·9678	3·9721	3·9763	3·9806	3·9848	3·9890	3·9931	3·9973	4·0014	4	8	13	17	21
16	4·0055	4·0096	4·0137	4·0178	4·0218	4·0259	4·0299	4·0339	4·0379	4·0419	4	8	12	16	20
17	4·0458	4·0498	4·0537	4·0576	4·0615	4·0654	4·0693	4·0731	4·0770	4·0808	4	8	12	16	19
18	4·0846	4·0884	4·0922	4·0960	4·0998	4·1035	4·1073	4·1110	4·1147	4·1184	4	8	11	15	19
19	4·1221	4·1258	4·1295	4·1331	4·1367	4·1404	4·1440	4·1476	4·1512	4·1548	4	7	11	15	18

20	4·1584	4·1619	4·1655	4·1690	4·1726	4·1761	4·1796	4·1831	4·1866	4·1901	4	7	11	14	18
21	4·1936	4·1970	4·2005	4·2039	4·2074	4·2108	4·2142	4·2176	4·2210	4·2244	3	7	10	14	17
22	4·2278	4·2312	4·2345	4·2379	4·2412	4·2446	4·2479	4·2512	4·2546	4·2579	3	7	10	13	17
23	4·2612	4·2644	4·2677	4·2710	4·2743	4·2775	4·2808	4·2840	4·2872	4·2905	3	7	10	13	16
24	4·2937	4·2969	4·3001	4·3033	4·3065	4·3097	4·3129	4·3160	4·3192	4·3224	3	6	10	13	16
25	4·3255	4·3287	4·3318	4·3349	4·3380	4·3412	4·3443	4·3474	4·3505	4·3536	3	6	9	12	16
26	4·3567	4·3597	4·3628	4·3659	4·3689	4·3720	4·3750	4·3781	4·3811	4·3842	3	6	9	12	15
27	4·3872	4·3902	4·3932	4·3962	4·3992	4·4022	4·4052	4·4082	4·4112	4·4142	3	6	9	12	15
28	4·4172	4·4201	4·4231	4·4260	4·4290	4·4319	4·4349	4·4378	4·4408	4·4437	3	6	9	12	15
29	4·4466	4·4495	4·4524	4·4554	4·4583	4·4612	4·4641	4·4670	4·4698	4·4727	3	6	9	12	14
30	4·4756	4·4785	4·4813	4·4842	4·4871	4·4899	4·4928	4·4956	4·4985	4·5013	3	6	9	11	14
31	4·5041	4·5070	4·5098	4·5126	4·5155	4·5183	4·5211	4·5239	4·5267	4·5295	3	6	8	11	14
32	4·5323	4·5351	4·5379	4·5407	4·5435	4·5462	4·5490	4·5518	4·5546	4·5573	3	6	8	11	14
33	4·5601	4·5628	4·5656	4·5684	4·5711	4·5739	4·5766	4·5793	4·5821	4·5848	3	5	8	11	14
34	4·5875	4·5903	4·5930	4·5957	4·5984	4·6011	4·6039	4·6066	4·6093	4·6120	3	5	8	11	14
35	4·6147	4·6174	4·6201	4·6228	4·6255	4·6281	4·6308	4·6335	4·6362	4·6389	3	5	8	11	13
36	4·6416	4·6442	4·6469	4·6495	4·6522	4·6549	4·6575	4·6602	4·6628	4·6655	3	5	8	11	13
37	4·6681	4·6708	4·6734	4·6761	4·6787	4·6814	4·6840	4·6866	4·6893	4·6919	3	5	8	11	13
38	4·6945	4·6971	4·6998	4·7024	4·7050	4·7076	4·7102	4·7129	4·7155	4·7181	3	5	8	10	13
39	4·7207	4·7233	4·7259	4·7285	4·7311	4·7337	4·7363	4·7389	4·7415	4·7441	3	5	8	10	13
40	4·7467	4·7492	4·7518	4·7544	4·7570	4·7596	4·7622	4·7647	4·7673	4·7699	3	5	8	10	13
41	4·7725	4·7750	4·7776	4·7802	4·7827	4·7853	4·7879	4·7904	4·7930	4·7955	3	5	8	10	13
42	4·7981	4·8007	4·8032	4·8058	4·8083	4·8109	4·8134	4·8160	4·8185	4·8211	3	5	8	10	13
43	4·8236	4·8262	4·8287	4·8313	4·8338	4·8363	4·8389	4·8414	4·8440	4·8465	3	5	8	10	13
44	4·8490	4·8516	4·8541	4·8566	4·8592	4·8617	4·8642	4·8668	4·8693	4·8718	3	5	8	10	13
45	4·8743	4·8769	4·8794	4·8819	4·8844	4·8870	4·8895	4·8920	4·8945	4·8970	3	5	8	10	13
46	4·8996	4·9021	4·9046	4·9071	4·9096	4·9122	4·9147	4·9172	4·9197	4·9222	3	5	8	10	13
47	4·9247	4·9272	4·9298	4·9323	4·9348	4·9373	4·9398	4·9423	4·9448	4·9473	3	5	8	10	13

TABLE 14.3. PROBITS—*continued*

	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1	2	3	4	5
48	4·9498	4·9524	4·9549	4·9574	4·9599	4·9624	4·9649	4·9674	4·9699	4·9724	3	5	8	10	13
49	4·9749	4·9774	4·9799	4·9825	4·9850	4·9875	4·9900	4·9925	4·9950	4·9975	3	5	8	10	13
50	5·0000	5·0025	5·0050	5·0075	5·0100	5·0125	5·0150	5·0175	5·0201	5·0226	3	5	8	10	13
51	5·0251	5·0276	5·0301	5·0326	5·0351	5·0376	5·0401	5·0426	5·0451	5·0476	3	5	8	10	13
52	5·0502	5·0527	5·0552	5·0577	5·0602	5·0627	5·0652	5·0677	5·0702	5·0728	3	5	8	10	13
53	5·0753	5·0778	5·0803	5·0828	5·0853	5·0878	5·0904	5·0929	5·0954	5·0979	3	5	8	10	13
54	5·1004	5·1030	5·1055	5·1080	5·1105	5·1130	5·1156	5·1181	5·1206	5·1231	3	5	8	10	13
55	5·1257	5·1282	5·1307	5·1332	5·1358	5·1383	5·1408	5·1434	5·1459	5·1484	3	5	8	10	13
56	5·1510	5·1535	5·1560	5·1586	5·1611	5·1637	5·1662	5·1687	5·1713	5·1738	3	5	8	10	13
57	5·1764	5·1789	5·1815	5·1840	5·1866	5·1891	5·1917	5·1942	5·1968	5·1993	3	5	8	10	13
58	5·2019	5·2045	5·2070	5·2096	5·2121	5·2147	5·2173	5·2198	5·2224	5·2250	3	5	8	10	13
59	5·2275	5·2301	5·2327	5·2353	5·2378	5·2404	5·2430	5·2456	5·2482	5·2508	3	5	8	10	13
60	5·2533	5·2559	5·2585	5·2611	5·2637	5·2663	5·2689	5·2715	5·2741	5·2767	3	5	8	10	13
61	5·2793	5·2819	5·2845	5·2871	5·2898	5·2924	5·2950	5·2976	5·3002	5·3029	3	5	8	10	13
62	5·3055	5·3081	5·3107	5·3134	5·3160	5·3186	5·3213	5·3239	5·3266	5·3292	3	5	8	11	13
63	5·3319	5·3345	5·3372	5·3398	5·3425	5·3451	5·3478	5·3505	5·3531	5·3558	3	5	8	11	13
64	5·3585	5·3611	5·3638	5·3665	5·3692	5·3719	5·3745	5·3772	5·3799	5·3826	3	5	8	11	13
65	5·3853	5·3880	5·3907	5·3934	5·3961	5·3989	5·4016	5·4043	5·4070	5·4097	3	5	8	11	14
66	5·4125	5·4152	5·4179	5·4207	5·4234	5·4261	5·4289	5·4316	5·4344	5·4372	3	5	8	11	14
67	5·4399	5·4427	5·4454	5·4482	5·4510	5·4538	5·4565	5·4593	5·4621	5·4649	3	6	8	11	14
68	5·4677	5·4705	5·4733	5·4761	5·4789	5·4817	5·4845	5·4874	5·4902	5·4930	3	6	8	11	14
69	5·4959	5·4987	5·5015	5·5044	5·5072	5·5101	5·5129	5·5158	5·5187	5·5215	3	6	9	11	14
70	5·5244	5·5273	5·5302	5·5330	5·5359	5·5388	5·5417	5·5446	5·5476	5·5505	3	6	9	12	14
71	5·5534	5·5563	5·5592	5·5622	5·5651	5·5681	5·5710	5·5740	5·5769	5·5799	3	6	9	12	15
72	5·5828	5·5858	5·5888	5·5918	5·5948	5·5978	5·6008	5·6038	5·6068	5·6098	3	6	9	12	15

73	5·6128	5·6158	5·6189	5·6219	5·6250	5·6280	5·6311	5·6341	5·6372	5·6403	3	6	9	12	15
74	5·6433	5·6464	5·6495	5·6526	5·6557	5·6588	5·6620	5·6651	5·6682	5·6713	3	6	9	12	16
75	5·6745	5·6776	5·6808	5·6840	5·6871	5·6903	5·6935	5·6967	5·6999	5·7031	3	6	10	13	16
76	5·7063	5·7095	5·7128	5·7160	5·7192	5·7225	5·7257	5·7290	5·7323	5·7356	3	7	10	13	16
77	5·7388	5·7421	5·7454	5·7488	5·7521	5·7554	5·7588	5·7621	5·7655	5·7688	3	7	10	13	17
78	5·7722	5·7756	5·7790	5·7824	5·7858	5·7892	5·7926	5·7961	5·7995	5·8030	3	7	10	14	17
79	5·8064	5·8099	5·8134	5·8169	5·8204	5·8239	5·8274	5·8310	5·8345	5·8381	4	7	11	14	18
80	5·8416	5·8452	5·8488	5·8524	5·8560	5·8596	5·8633	5·8669	5·8705	5·8742	4	7	11	14	18
81	5·8779	5·8816	5·8853	5·8890	5·8927	5·8965	5·9002	5·9040	5·9078	5·9116	4	7	11	15	19
82	5·9154	5·9192	5·9230	5·9269	5·9307	5·9346	5·9385	5·9424	5·9463	5·9502	4	8	12	15	19
83	5·9542	5·9581	5·9621	5·9661	5·9701	5·9741	5·9782	5·9822	5·8963	5·9904	4	8	12	16	20
84	5·9945	5·9986	6·0027	6·0069	6·0110	6·0152	6·0194	6·0237	6·0279	6·0322	4	8	13	17	21
85	6·0364	6·0407	6·0450	6·0494	6·0537	6·0581	6·0625	6·0669	6·0714	6·0758	4	9	13	18	22
86	6·0803	6·0848	6·0893	6·0939	6·0985	6·1031	6·1077	6·1123	6·1170	6·1217	5	9	14	18	23
87	6·1264	6·1311	6·1359	6·1407	6·1455	6·1503	6·1552	6·1601	6·1650	6·1700	5	10	15	19	23
88	6·1750	6·1800	6·1850	6·1901	6·1952	6·2004	6·2055	6·2107	6·2160	6·2212	5	10	15	21	26
89	6·2265	6·2319	6·2372	6·2426	6·2481	6·2536	6·2591	6·2646	6·2702	6·2759	5	11	16	22	27
90	6·2816	6·2873	6·2930	6·2988	6·3047	6·3106	6·3165	6·3225	6·3285	6·3346	6	12	18	24	29
91	6·3408	6·3469	6·3532	6·3595	6·3658	6·3722	6·3787	6·3852	6·3917	6·3984	6	13	19	26	32
92	6·4051	6·4118	6·4187	6·4255	6·4325	6·4395	6·4466	6·4538	6·4611	6·4684	7	14	21	28	35
93	6·4758	6·4833	6·4909	6·4985	6·5063	6·5141	6·5220	6·5301	6·5382	6·5464	8	16	24	31	39
94	6·5548	6·5632	6·5718	6·5805	6·5893	6·5982	6·6072	6·6164	6·6258	6·6352	9	18	27	36	45
95	6·6449	6·6546	6·6646	6·6747	6·6849	6·6954	6·7060	6·7169	6·7279	6·7392					
	97	100	101	102	105	106	109	110	113	115					
96	6·7507	6·7624	6·7744	6·7866	6·7991	6·8119	6·8250	6·8384	6·8522	6·8663					
	117	120	122	125	128	131	134	138	141	145					
97	6·8808	6·8957	6·9110	6·9268	6·9431	6·9600	6·9774	6·9954	7·0141	7·0335					
	149	153	158	163	169	174	180	187	194	202					

TABLE 14.3. PROBITS—*continued*

	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09	1	2	3	4	5
98·0	7·0537	7·0558	7·0579	7·0600	7·0621	7·0642	7·0663	7·0684	7·0706	7·0727	2	4	6	8	11
98·1	7·0749	7·0770	7·0792	7·0814	7·0836	7·0858	7·0880	7·0902	7·0924	7·0947	2	4	7	9	11
98·2	7·0969	7·0992	7·1015	7·1038	7·1061	7·1084	7·1107	7·1130	7·1154	7·1177	2	5	7	9	12
98·3	7·1201	7·1224	7·1248	7·1272	7·1297	7·1321	7·1345	7·1370	7·1394	7·1419	2	5	7	10	12
98·4	7·1444	7·1469	7·1494	7·1520	7·1545	7·1571	7·1596	7·1622	7·1648	7·1675	3	5	8	10	13
98·5	7·1701	7·1727	7·1754	7·1781	7·1808	7·1835	7·1862	7·1890	7·1917	7·1945	3	5	8	11	14
98·6	7·1973	7·2001	7·2029	7·2058	7·2086	7·2115	7·2144	7·2173	7·2203	7·2232	3	6	9	12	14
98·7	7·2262	7·2292	7·2322	7·2353	7·2383	7·2414	7·2445	7·2476	7·2508	7·2539	3	6	9	12	15
98·8	7·2571	7·2603	7·2636	7·2668	7·2701	7·2734	7·2768	7·2801	7·2835	7·2869	3	7	10	13	17
98·9	7·2904	7·2938	7·2973	7·3009	7·3044	7·3080	7·3116	7·3152	7·3189	7·3226	4	7	11	14	18
99·0	7·3263	7·3301	7·3339	7·3378	7·3416	7·3455	7·3495	7·3535	7·3575	7·3615	4	8	12	16	20
99·1	7·3656	7·3698	7·3739	7·3781	7·3824	7·3867	7·3911	7·3954	7·3999	7·4044	4	9	13	17	22
99·2	7·4089	7·4135	7·4181	7·4228	7·4276	7·4324	7·4372	7·4422	7·4471	7·4522	5	10	14	19	24
99·3	7·4573	7·4624	7·4677	7·4730	7·4783	7·4838	7·4893	7·4949	7·5006	7·5063	5	11	16	22	27
99·4	7·5121	7·5181	7·5241	7·5302	7·5364	7·5427	7·5491	7·5556	7·5622	7·5690	6	13	19	25	32
99·5	7·5758	7·5828	7·5899	7·5972	7·6045	7·6121	7·6197	7·6276	7·6356	7·6437					
99·6	7·6521	7·6606	7·6693	7·6783	7·6874	7·6968	7·7065	7·7164	7·7266	7·7370					
99·7	7·7478	7·7589	7·7703	7·7822	7·7944	7·8070	7·8202	7·8338	7·8480	7·8627					
99·8	7·8782	7·8943	7·9112	7·9290	7·9478	7·9677	7·9889	8·0115	8·0357	8·0618					
99·9	8·0902	8·1214	8·1559	8·1947	8·2389	8·2905	8·3528	8·4316	8·5401	8·7190					

The probit corresponding to a given percentage is the normal deviate (increased by 5 to avoid negative values) exceeded by this percentage of the population.

Table 14.3 is reprinted from Table IX of Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Ltd., Edinburgh) by permission of the Authors and Publishers.

TABLE 14.4

CONSTANTS FOR DETERMINING THE CORRECTED PROBIT AND THE WEIGHTING COEFFICIENT FROM THE EXPECTED PROBIT, Y. CORRECTED PROBITS FOR 100 AND 0 % MORTALITY, RESPECTIVELY, ARE GIVEN DIRECTLY UNDER THE HEADINGS "MAXIMUM CORRECTED PROBIT" AND "MINIMUM CORRECTED PROBIT."

From Bliss, *Quart. J. Pharm. Pharmacol.*, **11**, 192, 1938.

Expected probit Y	Maximum corrected probit $Y+Q/Z$	Range $1/Z$	Minimum corrected probit $Y-P/Z$	Weighting coefficient Z^2/PQ	Maximum corrected probit $Y+Q/Z$	Range $1/Z$	Minimum corrected probit $Y-P/Z$	Expected probit Y
5·0	6·253	2·507	3·747	·6366	6·253	2·507	3·747	5·0
5·1	6·259	2·519	3·740	·6343	6·260	2·519	3·741	4·9
5·2	6·276	2·557	3·719	·6274	6·281	2·557	3·724	4·8
5·3	6·302	2·622	3·680	·6161	6·320	2·622	3·698	4·7
5·4	6·336	2·715	3·620	·6005	6·380	2·715	3·664	4·6
5·5	6·376	2·840	3·536	·5810	6·464	2·840	3·624	4·5
5·6	6·423	3·001	3·422	·5579	6·578	3·001	3·577	4·4
5·7	6·475	3·203	3·272	·5316	6·728	3·203	3·525	4·3
5·8	6·531	3·452	3·079	·5026	6·921	3·452	3·469	4·2
5·9	6·592	3·758	2·834	·4714	7·166	3·758	3·408	4·1
6·0	6·656	4·133	2·523	·4386	7·477	4·133	3·344	4·0
6·1	6·723	4·590	2·132	·4047	7·867	4·590	3·277	3·9
6·2	6·793	5·150	1·643	·3703	8·357	5·150	3·207	3·8
6·3	6·865	5·835	1·030	·3359	8·970	5·835	3·135	3·7
6·4	6·939	6·679	0·261	·3020	9·739	6·679	3·061	3·6
6·5	7·016	7·721	—	·2691	—	7·721	2·984	3·5
6·6	7·094	9·015	—	·2375	—	9·015	2·906	3·4
6·7	7·174	10·633	—	·2077	—	10·633	2·826	3·3
6·8	7·255	12·666	—	·1799	—	12·666	2·745	3·2
6·9	7·338	15·240	—	·1544	—	15·240	2·662	3·1
7·0	7·421	18·522	—	·1311	—	18·522	2·579	3·0
7·1	7·506	22·736	—	·1103	—	22·736	2·494	2·9
7·2	7·592	28·189	—	·0918	—	28·189	2·408	2·8
7·3	7·679	35·302	—	·0756	—	35·302	2·321	2·7
7·4	7·766	44·654	—	·0617	—	44·654	2·234	2·6
7·5	7·854	57·05	—	·0498	—	57·05	2·146	2·5
7·6	7·943	73·62	—	·0398	—	73·62	2·057	2·4
7·7	8·033	95·96	—	·0314	—	95·96	1·967	2·3
7·8	8·123	126·34	—	·0246	—	126·34	1·877	2·2
7·9	8·213	168·00	—	·0190	—	168·00	1·787	2·1
8·0	8·305	225·6	—	·0146	—	225·6	1·695	2·0
8·1	8·396	306·1	—	·0110	—	306·1	1·604	1·9
8·2	8·488	419·4	—	·0083	—	419·4	1·512	1·8
8·3	8·581	580·5	—	·0061	—	580·5	1·419	1·7
8·4	8·673	811·5	—	·0045	—	811·5	1·327	1·6

In calculating the first approximation to the dosage-response line, it is sufficiently accurate to determine the expected probits to the first decimal place, but in further approximations, if these are found to be necessary, it is best to calculate expected probits to the second or even the third decimal place and to interpolate in Table 14.4 to obtain the corresponding values of the corrected probits and range. The easiest method of interpolation may be illustrated by supposing that we have an expected probit of 6·031. We then read off the maximum corrected probit corresponding to an expected probit of 6·0 and multiply it by 0·69, read off the maximum corrected probit corresponding to an expected probit of 6·1 and multiply it by 0·31 and add the results together. This will be the maximum corrected probit for an expected probit of 6·031. The interpolated values of the range and the weighting coefficient are determined in an exactly similar manner. The weighting coefficient corresponding to a corrected probit is the value of w in the same row of the table, multiplied by the number of observations in the group. When the empirical probit differs from the expected probit by less than 0·05, particularly in the range from 4 to 6 probits, the empirical value will agree sufficiently well with the corrected probit for most purposes of calculation. In calculations aiming at extreme accuracy, however, corrected probits should always be used.

14.7. A numerical example

Writing the equation for the corrected probit:

$$Y_c = \left(Y + \frac{Q}{Z}\right) - \frac{q}{Z}$$

we use Table 14.4 by reading directly from the Table the value of the maximum corrected probit $\left(Y + \frac{Q}{Z}\right)$ and from it we subtract q times the range, where it will be remembered q is the observed proportion of *non-reactors*. For expected values of less than 5 we read off the minimum corrected probit in the next to last column and add to it the range multiplied by p, the *proportion reacting*, according to the complementary equation:

$$Y_c = \left(Y - \frac{P}{Z}\right) + \frac{p}{Z}$$

No correction for the range is necessary in the case of 0 or 100%

response. With an expected probit of 5·5, and an observed proportion, 74%, reacting (empirical probit=5·643):

$$Y_t = 6{\cdot}376 - 0{\cdot}26 \times 2{\cdot}840 = 5{\cdot}538$$

with a weight factor of $0{\cdot}5810 n_p$. It may be noted that the corrected probit is not necessarily nearer to the expected probit than is the empirical probit, but that it usually is.

This rather involved procedure for the computation of corrected probits is necessary because we are dealing with a discontinuous variate, p, the gradations of which may be very coarse when a small number of animals is used. The computational procedure is concerned with smoothing the effect of this coarse gradation and relating it to the continuous variable, Q.

14.8. Correction for reactions in the controls

When untreated control groups contain a proportion of test objects which exhibit the characteristic response, it is necessary to correct the observed responses in treated groups. If p_c, p_o and p are respectively the observed control percentage response, the observed experimental percentage response in any given group and the adjusted percentage response, then:

$$p = \frac{100(p_o - p_c)}{100 - p_c}$$

It is unsatisfactory to conduct assays by a technique involving control groups with many responses, but sometimes unavoidable. In such cases, weighting coefficients and certain other parts of the computational procedure need modification. This has recently been fully described by Finney (*Probit Analysis*, Cambridge University Press, 1947) and will not be further discussed here. The need for its use with other than toxicological data is rare.

CHAPTER 15

CALCULATIONS INVOLVING PROBITS

15.1. Unbalanced data

As with other material, we can simplify calculations involving probits by a suitable choice of doses and of the numbers of animals used in a group. The calculation is, nevertheless, always more involved than when dealing with correspondingly simplified data for continuous responses. There is always the weight factor by which allowance is made in computations for inequalities in the variance, despite our using the same number of animals per group, and the calculation of corrected probits further increases the amount of work involved. When doses are unequally spaced on a logarithmic scale and the number of animals varies from group to group, computational procedure is correspondingly more complex and all the remarks previously made about simplification in design are applicable here.

We shall get the worst over by computing a line, as an example, from a test containing unequal spacing of doses and varying numbers of animals per group. This is sometimes unavoidable, but rarely so, and the reader must be familiar with the computational procedure in case he needs it. If it is anticipated that some test objects will be lost in the course of the test, this can sometimes be provided for by putting extra members in each group and selecting strictly at random from the remainder, if an excess still remains in some groups. The procedure must be used with extreme caution, as it would bias tests if applied uncritically in circumstances where loss of a test object may be correlated with the presence or absence of a response. Table 15.1 lists the protocols of a test, designed for establishing a dose-response line, in which groups of spayed female rats received doses of 1 to 6 international units of International Standard oestrone. The characteristic response used in this investigation was the disappearance of leucocytes from the vaginal smear after injection of the drug. Since there is no regularity in the spacing of the log doses, we take logarithms to base 10 and list them against the dose in units in the Table; next we list the number of animals per group and the number reacting and from

this derive the percentage of reactors, which varies from 0 to 100. Empirical probits are then read off from Table 14.3 for the four intermediate dosage groups and a provisional line is fitted, the dotted line in Figure 15.1. This provisional line has an approximate

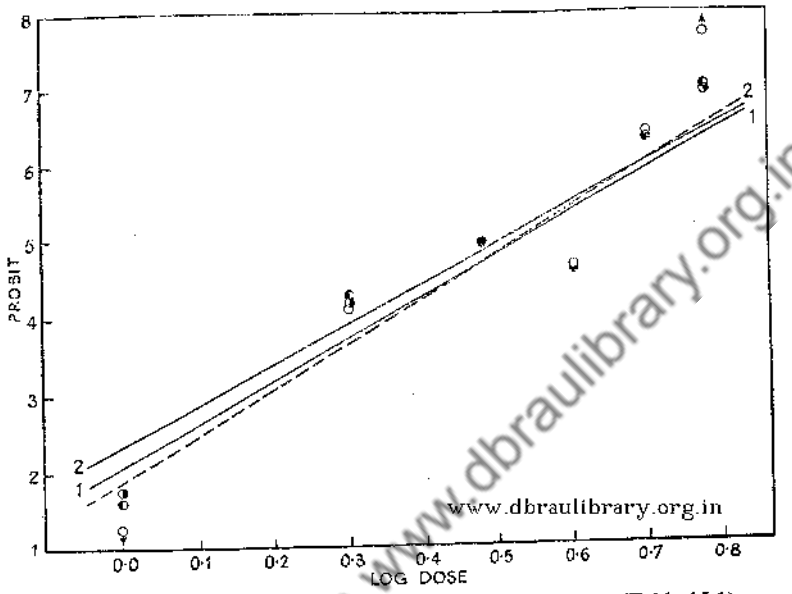


FIG. 15.1. Fitting a dose-response line to enumeration data (Table 15.1).
○ - - - - ○, empirical probits and preliminary line.
◐ 1 —— 1, corrected probits and first approximation.
◑ 2 —— 2, corrected probits and second approximation.

slope of 5·90. In plotting a provisional line we may conveniently use log graph paper, so that we can plot the actual dose directly against the probit.

TABLE 15.1

BASIC DATA FOR THE ESTABLISHMENT OF A QUANTAL LOG DOSE-RESPONSE LINE

Dose in units	Log_{10} dose	No. of animals	No. reacting	Percentage reacting	Empirical probit
1	0·000	25	0	0·0	—
2	0·301	11	2	18·2	4·092
3	0·477	27	13	48·1	4·952
4	0·602	19	7	36·8	4·663
5	0·699	12	11	91·7	6·385
6	0·778	17	17	100·0	—

Provisional $b = 5{\cdot}90$.

TABLE 15.2

ESTIMATION OF A QUANTAL LOG DOSE-RESPONSE LINE FOR THE DATA OF TABLE 15.1 (FIRST APPROXIMATION)

Log dose	No. of animals	Expected probit	Corrected probit	Weight					
$\bar{X}_p$	n_p	Y	Y_c	$n_p w_p$	$n_p w_p \bar{X}_p$	$n_p w_p Y_c$	$n_p w_p \bar{X}_p Y_c$	$n_p w_p \bar{X}_p^2$	$n_p w_p Y_c^2$
0·000	25	1·9	1·604	0·275	0·0000	0·4411	0·0000	0·0000	0·7075
0·301	11	3·6	4·277	3·322	0·9999	14·2082	4·2766	0·3010	60·7685
0·477	27	4·7	4·954	16·632	7·9335	82·3949	39·3026	3·7843	408·1843
0·602	19	5·4	4·619	11·419	6·9742	52·7444	32·2138	4·1985	243·6264
0·699	12	6·0	6·313	5·268	3·6823	33·2569	23·2464	2·5739	209·9508
0·778	17	6·5	7·016	4·573	3·5578	32·0842	24·9165	2·7680	225·1027
			Totals	41·489	23·1477	215·1297	124·0009	13·6257	1,148·3402

15.2. The first approximation to the dose-response line

We now proceed to determine the quantities listed in each column of Table 15.2, writing down in the first two columns the logarithm of the dose and the number of animals for convenience in computation. Expected probits, as read back from the provisional line, are listed in the third column of Table 15.2 and from them, as described in the preceding chapter, we compute corrected probits and the weight of each observation; the last five columns in the Table are then computed for the determination of the weighted sums of $\bar{X}_p$, Y_c and of their squares and products. The last column of all, under the heading $n_p w_p Y_c^2$, is not needed in the calculation of the dose-response line, but will be needed when we test for homogeneity. Columns 5 to 10 are summed and the weighted means $\bar{X}$ and $\bar{Y}_c$ computed by dividing the sums of columns 6 and 7 respectively by the sum of the weights. Thus:

$$\bar{X}=\frac{23{\cdot}1477}{41{\cdot}489}=0{\cdot}557924, \text{ and } \bar{Y}_c=\frac{215{\cdot}1297}{41{\cdot}489}=5{\cdot}185223$$

The weighted sums of squares and products of the deviations from the two means are:

$$Sn_p w_p \bar{x}_p^2 = Sn_p w_p \bar{X}_p^2 - \bar{X} Sn_p w_p \bar{X}_p = 13{\cdot}6257 - 12{\cdot}9147 = 0{\cdot}7110$$

$$Sn_p w_p y_c^2 = Sn_p w_p Y_c^2 - \bar{Y}_c Sn_p w_p Y_c = 1{,}148{\cdot}3402 - 1{,}115{\cdot}4955 = 32{\cdot}8447$$

$$Sn_p w_p \bar{x}_p y_c = Sn_p w_p \bar{X}_p Y_c - \bar{Y}_c Sn_p w_p \bar{X}_p = 124{\cdot}0009 - 120{\cdot}0260 = 3{\cdot}9749$$

When using this method of computation, calculate two more decimal places than are required in the final sums of squares and products. In this instance, the dropping of one decimal place would not seriously affect the accuracy of the result.

The slope b, which is $\dfrac{Sn_p w_p \bar{x}_p y_c}{Sn_p w_p \bar{x}_p^2}=\dfrac{3{\cdot}9749}{0{\cdot}7110}=5{\cdot}5906$. The dose-response line is then determined by substituting the quantities just calculated in the formula:

$$E=\bar{Y}_c+b(X-\bar{X})$$

$$\text{whence } E=5{\cdot}1852+5{\cdot}5906(X-0{\cdot}5579)$$

$$\text{or } E=2{\cdot}0661+5{\cdot}59X.$$

15.3. Goodness of fit

We compute χ^2 for estimating how well the data are fitted by an analogous method to isolating the sum of squares attributable to departures from linear regression in previous assays:

$$\chi^2=Sn_p w_p y_c^2 - b(Sn_p w_p \bar{x}_p y_c)$$

In this example:

$$\chi^2 = 32{\cdot}84 - 22{\cdot}22 = 10{\cdot}62$$

The significance of χ^2 depends on the number of degrees of freedom in the experiment. This number is two less than the number of original percentages from which the curve has been plotted, since two constants have been computed from the observations. Bliss (*Ann. Appl. Biol.*, **22**, 134, 1935) has suggested that this conventional rule for the number of degrees of freedom is not strictly applicable when some of the doses have elicited 0 or 100% response, for the corrected probits corresponding to these responses have been determined primarily from the evidence of the remaining observations. An approximation to the "true" number of degrees of freedom which he suggests is to combine groups showing 0 or 100% response with those next to them, if necessary at both ends of the curve, until the sum of the expected actual number of reactors, or non-reactors as the case may be, exceeds one individual. In this example the expected percentage response at the lowest dose is 0·17 and on the highest dose is 92·2 and we should thus expect $0{\cdot}17 \times 25 = 0{\cdot}043$ and $7{\cdot}8 \times 17 = 1{\cdot}33$ individuals in these groups to react and not to react respectively. The group on the lowest dose, when combined with that on the next higher dose, gives an expectation of 1·20 individuals and the two together are counted as contributing only one degree of freedom. We are thus dealing with the equivalent of five groups and the number of degrees of freedom is $5-2=3$. We therefore enter the Table of χ^2 with three degrees of freedom and find that a value of χ^2 as high as 10·62 should be expected in between 1% and 2% of cases ($P = 0{\cdot}02 - 0{\cdot}01$), and thus the series does not test as homogeneous.

It is difficult to justify this procedure and a more exact evaluation of χ^2 may be made in these circumstances by the "longhand" method of calculating, from the regression equation, expected probits and from them the expected percentages of reactions. Each of these in turn multiplied by the corresponding n_p and divided by 100 gives the expected numbers of reactions.

Then $$\chi^2 = S\frac{(\textit{number reacting minus expected number})^2}{P(1-P)n_p},$$

where P is the proportion of expected reactions. Groups in which nearly 100% or 0% of reactions are *expected* should be separately combined with the next group(s) at each end of the dose-response line as in a usual χ^2 test with small expected numbers.

Finney (*Probit Analysis*, Cambridge University Press, 1947) finds that in groups of up to 10 test objects it is probably safe to have not less than two expected reactions, but in groups of 30 or more there should be the conventional number of five expected reactions. There will be two less degrees of freedom than there are groups remaining after these combinations have been made. In the combined groups, each contribution to χ^2 is given by:

$$\chi^2 = \frac{Sn_p \times (Number\ reacting\ minus\ expected\ number)^2}{Expected\ number\ reacting \times Expected\ number\ not\ reacting}$$

The calculation for the present example is shown in Table 15.3, in which the first three groups and the last two groups are combined. The new $\chi^2_{[1]}$ is 10·45, and since only one degree of freedom now exists there is no doubt of the heterogeneity. The number of degrees of freedom with which the χ^2 table is entered is indicated by the bracketed subscript to χ^2.

TABLE 15.3

CALCULATION OF χ^2 FOR THE DATA OF TABLE 15.2

Dose in units	Number reacting		Expected reactors		Difference	χ^2
1	0		0·04			
2	2	15	1·16	11·88	3·12	1·01
3	13		10·68			
4	7		12·67		5·67	7·62
5	11		10·02			
6	17	28	15·67	25·69	2·31	1·82
					Sum	10·45

15.4. Second approximation to the dose-response line

In view of the lack of agreement between the provisional slope (5·90) and the slope of the first approximation to the dose-response line (5·59), and also in view of the bad fit as demonstrated by χ^2, we must compute a second approximation. The results are shown in Table 15.4.

Expected probits have been calculated from the first approximation to the third place of decimals, although this is a rather unnecessary refinement in the present instance, and the corrected probits and weights determined by interpolation. In view of the

TABLE 15.4

SECOND APPROXIMATION TO THE DOSE-RESPONSE LINE FOR THE DATA OF TABLE 15.1

Log dose, X_p	Expected probit, Y	Corrected probit, Y_c	Weight $n_p w_p$	$n_p w_p X_p$	$n_p w_p Y_c$	$n_p w_p X_p Y_c$	$n_p w_p X_p^2$	$n_p w_p Y_c^2$
0·000	2·066	1·756	0·438	0·00000	0·76913	0·00000	0·00000	1·35059
0·301	3·749	4·171	3·881	1·16818	16·18765	4·87248	0·35162	67·51869
0·477	4·733	4·957	16·735	7·98260	82·95540	39·56973	3·80770	411·20992
0·602	5·432	4·608	11·290	6·79658	52·02432	31·31864	4·09154	239·72807
0·699	5·974	6·306	5·365	3·75014	33·83169	23·64835	2·62135	213·34264
0·778	6·416	6·951	5·044	3·92423	35·06084	27·27733	3·05305	243·70790
		Totals	42·753	23·62173	220·82903	126·68653	13·92526	1,176·85781

small numbers of observations involved, the calculations in Table 15.4 have in fact been carried to more decimal places than strictly necessary. When working with a machine it doesn't much matter, and these places were carried for safety. Without a machine, however, the carrying of these extra figures would mean unnecessary labour. Computational procedure is exactly as before and we arrive at the following values for the various statistics:

$$\bar{X}=\frac{23{\cdot}62173}{42{\cdot}753}=0{\cdot}5525163$$

$$\bar{Y}_c=\frac{220{\cdot}82903}{42{\cdot}753}=5{\cdot}1652289$$

$$Sn_pw_p\bar{x}_p^2=13{\cdot}92526-13{\cdot}05139=0{\cdot}87387$$

$$Sn_pw_py_c^2=1{,}176{\cdot}85781-1{,}140{\cdot}63249=36{\cdot}22532$$

$$Sn_pw_p\bar{x}_py_c^2=126{\cdot}68653-122{\cdot}01164=4{\cdot}67489$$

$$b=\frac{4{\cdot}67489}{0{\cdot}87387}=5{\cdot}34964$$

$$\chi^2_{[3]}=36{\cdot}2253-25{\cdot}0090=11{\cdot}22,$$

and with three degrees of freedom, P lies between 0·01 and 0·02.

The value of $\chi^2_{[1]}$ calculated by the method of Table 15.3 is 10·86.

The new line is estimated as:

$$E=5{\cdot}1652+5{\cdot}3496(X-0{\cdot}55252)$$
$$=2{\cdot}209+5{\cdot}350X.$$

Our new values differ quite appreciably from those previously obtained, but χ^2 still shows as bad a fit. If this were not the case it might be worth fitting a third approximation (but if χ^2 did not indicate significant departures from the computed line the two successive approximations would probably have agreed better). It is now clear that the series of observations is not well fitted by a linear probit-log dose relationship and there is no point in trying for a more exact line. Examination of the general trend of the results does not indicate that the departure from linearity is of a systematic type, but merely that the scatter of the points about the line is too great for us to suppose that we have sampled a homogeneous population.

15.5. The estimation of errors

If χ^2 does not indicate a greater variation between the observations and the computed curve than is to be expected by chance,

i.e. if the variation is within the limits for $P=0{\cdot}05$, we may compute the following variances as:

$$V\bar{Y}_c=\frac{1}{Sn_pw_p}$$

$$Vb=\frac{1}{Sn_pw_p\bar{x}_p{}^2}$$

When χ^2 indicates a significant departure from homogeneity and these statistics are not really an adequate description of the data, the variances are:

$$V\bar{Y}_c=\frac{\chi^2}{kSn_pw_p}$$

and $$Vb=\frac{\chi^2}{kSn_pw_p\bar{x}_p{}^2}$$

where k equals the number of degrees of freedom for the χ^2 test. This follows from the fact that the average value of $\frac{\chi^2}{k}$ in sampling is unity if the normal hypothesis is true, but when χ^2 shows significant departures from the line, we cannot hold to the normal hypothesis. Our estimates of each variance are then multiplied by the "*heterogeneity factor*" which is the fraction $\frac{\chi^2}{k}$.

15.6. The median effective dose

In describing the results of tests to determine the dose-response relationships and relative potencies when the response is quantal, it is usual to determine the *median effective dose* (M.E.D.). This is the dose to which 50% of test objects react and therefore give a probit of 5·0. The median effective dose is computed by substituting 5 for E in the equation above, whence M.E.D.=antilog 0·5217. A test designed for the computation of the median effective dose gains in accuracy as $\bar{Y}$ approaches 5. The variance of the median effective dose may be computed from the general equation for the average variance of log dose X corresponding to a given probit, Y:

$$VX=\frac{Vb(Y-\bar{Y})^2+(V\bar{Y})b^2}{b^4}$$

This equation can be used when the data are not homogeneous, as decided by the χ^2 test, as well as when they are.

In the present example, where χ^2 is based on so few degrees of freedom, the new variances are practically useless, as they must be

used in conjunction with the statistic t, and would give such wide limits of error that almost no information about the effectiveness of the treatment is in fact available. In the absence of other tests of a like nature from which an overall estimate of the value of the heterogeneity factor could be formed, we should conclude that a repeat test is indicated. If χ^2 did not indicate significant heterogeneity, we should estimate

$$V\bar{Y}_c = \frac{1}{42{\cdot}753} = 0{\cdot}02339$$

$$Vb = \frac{1}{0{\cdot}87387} = 1{\cdot}144$$

When
$$VX = \frac{1{\cdot}144(0{\cdot}1652)^2 + 0{\cdot}02339 \times 28{\cdot}62}{819{\cdot}28}$$
$$= 0{\cdot}0008551$$

The approximate limits of error for $M(P-0{\cdot}05)$ are then:

$$M \pm \sqrt{0{\cdot}0008551}t, \text{ where } t = 1{\cdot}960$$
or 0·4643 to 0·5789,

corresponding to 2·91 to 3·79 units as the limits for the M.E.D.

When allowance is made for heterogeneity, we multiply the variances by 3·740 for the Bliss approximation, and by 10·86 for the more exact evaluation of $\chi^2_{[1]}$; moreover, we use t corresponding to three and one degrees of freedom respectively and arrive at the limits:

Bliss method: 2·20—5·03 units
$\chi^2_{[1]} = 10{\cdot}86$: 0·20—55·8 units,

the M.E.D. itself being estimated throughout as 3·32 units. These are still approximate limits. The fiducial limits of error of the M.E.D. would be still wider, since Vb is relatively large, but since these are not usually of practical interest except in actual assays, discussion is postponed to the next Chapter.

CHAPTER 16

PROBIT ASSAYS

16.1. The design of assays

Since the variance of a probit is not constant at all levels of response, the analysis of variance is not directly applicable, particularly when responses give very high or very low percentages. However, when dealing with quantal responses we use a theoretical estimate of the variance unless the χ^2 test indicates significant departure from homogeneity; therefore the performance of an analysis of variance after the style used in assays involving graded responses is not necessary.

The normal use of the analysis in assays of potency is to reduce the error mean square by eliminating irrelevant sources of variation, and since in assays using graded responses there is no theoretical estimate available of the mean square for error, the analysis performs a very necessary function. This does not mean that when the response is quantal proper randomisation and the use of certain restrictions in the design of the test should not be carried out. On the contrary, by using one of the restricted randomisation designs from which the maximum of information may be extracted by a variance analysis in other cases, we ensure as far as is possible that the test for homogeneity by means of χ^2 will show a minimal departure from a linear dose-response line if, indeed, this line adequately describes the relationship between log dose and response. We should not, however, introduce unnecessary sources of variation the effects of which we cannot eliminate, and thus would not distribute, for instance, different strains of animals between dosage groups unless this were unavoidable. If we had to do so, as when dealing with litters of animals, we should nevertheless follow the usual plan and allot one animal of each litter to each dose, for we should thus help to guarantee that any extra variation introduced would tend to lower the value of b, with minimal elevation in the value of χ^2, and should still give a satisfactory assay. Methods for the statistical handling of such tests so as to eliminate differences between litters have yet to be described.

16.2. Computational procedure

(a) If the standard and unknown may well differ in slope.

The computation for an assay using probits follows the same method as in Chapter 15. Weighted means of X and Y_c and of their sums of squares and products are computed for each substance separately, by fitting separate provisional lines for the standard and the unknown and computing expected and then corrected probits by the usual method. In a well-designed assay these provisional lines need not be fitted by eye, but can more accurately be computed by the following simple method.

The unweighted sum of the empirical probits for each substance is divided by the number of doses and this gives an unweighted value of $\bar{Y}$; a similar unweighted value of $\bar{X}$ is also computed and an unweighted value of b is obtained by the following formula:

1. If there are two doses of the standard and unknown:

$$b=\frac{\bar{Y}_1-\bar{Y}_2}{\bar{X}_1-\bar{X}_2}$$

where $\bar{Y}_1$ and $\bar{Y}_2$ and $\bar{X}_1$ and $\bar{X}_2$ are the responses and log doses for the two doses for each substance respectively.

2. If there are three doses of each substance:

$$b=\frac{\bar{Y}_1-\bar{Y}_3}{\bar{X}_1-\bar{X}_3}$$

3. If there are four doses of each substance:

$$b=\frac{3(\bar{Y}_1-\bar{Y}_4)+(\bar{Y}_3-\bar{Y}_2)}{10I}$$

where I is the interval between log doses.

A provisional line may thus be calculated for each substance and the expected probits obtained by computation instead of by graphical fitting. From these values the first approximation is calculated and a second approximation follows by the usual technique if it seems indicated. With a good initial fit the second approximation is often unnecessary. When the line to be finally used has been computed for each substance, we test the value of χ^2 for each separate value of b and for the difference between the two values of b where:

$$\chi_s^2=S_s n_p w_p y_c^2-b_s S_s n_p w_p \bar{x}_p y_c$$

for the standard and similarly for the unknown. The added values of these χ^2s should be tested with four degrees of freedom

less than the *total* number of dosage groups. The value of χ^2 for the difference between slopes is given by the formula:

$$\chi_{\text{diff}}^2 = \frac{(b_s - b_u)^2}{\frac{1}{S_s n_p w_p \bar{x}_p^2} + \frac{1}{S_u n_p w_p \bar{x}_p^2}}$$

With this value of χ^2 is associated one degree of freedom. Hence it should not exceed 3·84 if there is no difference between the two values of the slope at the 5 % level of significance. If no significance is to be attached to the value of χ^2 it means that the samples are qualitatively similar and that one combined estimate of the slope may validly be employed. Otherwise, of course, the assay is not valid.

The value of the first approximation to the combined slope, b_c, is such that:

$$b_c = \frac{S_s n_p w_p \bar{x}_p y_c + S_u n_p w_p \bar{x}_p y_c}{S_s n_p w_p \bar{x}_p^2 + S_u n_p w_p \bar{x}_p^2}$$

(*b*) *If it is probable or certain that a common slope is valid.*

It may be clear by inspection that a common slope for standard and unknown can be used, whereupon the above process can be shortened and a further cycle of calculations probably avoided by plotting provisional lines which are parallel initially. These lines will have a common slope estimated by combining the unweighted values described above. The χ^2 calculations can similarly be run together, this method of calculation being particularly useful when several unknowns are being simultaneously assayed against a standard.

In the general case, we calculate:

$$\text{Total } \chi^2 = SSn_p w_p y_c^2 - b_c SSn_p w_p \bar{x}_p y_c$$

If there are k dosage groups per substance and N substances in all, this total χ^2 is associated with $S(k-1)-1$ degrees of freedom, or $N(k-1)-1$ if there are equal numbers of doses of all substances. We also calculate a similar χ^2 separately for each substance, including the standard, and obtain N values for χ_s^2 and χ_u^2s. With each of these is associated $k-2$ degrees of freedom. By subtraction, the following analysis of χ^2 may be constructed:

Source of variation	Degrees of freedom	Sum of squares, etc.
Parallelism	$N-1$	By subtraction
Heterogeneity	$S(k-2)$	From sum of individual χ^2s as above
Total	$S(k-1)-1$	As above

If the χ^2 for heterogeneity is not significantly large, we test the significance of the χ^2 for parallelism with $N-1$ degrees of freedom, but if there is significant heterogeneity we determine mean squares by dividing each χ^2 by the relevant degrees of freedom and test these by the F-test. In the latter case, the heterogeneity factor must enter all further calculations as in the example of the preceding Chapter.

The reciprocal of b_c, $\frac{1}{b_c}$, sometimes termed λ, is the standard deviation for the population of the minimum effective log dose. M, the log ratio of potency, is then given by the equation:

$$M=\bar{X}_s-\bar{X}_u-\frac{1}{b_c}(\bar{Y}_s-\bar{Y}_u)$$

The variance of b_c is given by:

$$Vb_c=\frac{1}{S_s n_p w_p \bar{x}_p^2+S_u n_p w_p \bar{x}_p^2}$$

The variance of M is then such that:

$$VM=\frac{1}{b_c^2}\left(\frac{1}{S_s n_p w_p}+\frac{1}{S_u n_p w_p}+\frac{(\bar{Y}_s-\bar{Y}_u)^2 Vb_c}{b^2}\right)$$

This, as before, is not the formula giving fiducial limits and should be used only if $\frac{b_c}{s_{b_c}}$ is relatively large. Fiducial limits may be calculated exactly as for other types of assay, but when calculating fiducial limits for quantal assays t is normally distributed and the value used is therefore equal to that in the t Table for an infinite number of degrees of freedom. The calculation is described in 16.4.

16.3. An example using balanced dosage groups

Table 16.1 and Figure 16.1 illustrate the computation in the comparison of two yeast extracts as curative agents in pigeons deficient in vitamin B_1 (Burn, *Biological Standardisation*, C.U.P., 1937). The birds are prepared for the test by being fed on a diet of polished rice and water. After a few weeks many of the birds develop a neuritis, the indication of which is retraction of the head and spontaneous convulsions. They are then ready for the test. Groups of birds with persistent retraction of the head are randomised and given the vitamin preparation by mouth and are scored as cured if the symptoms disappear for at least one day.

TABLE 16.1

THE COMPARISON OF TWO YEAST EXTRACTS AS ANTINEURITIC AGENTS IN PIGEONS

Extract	Dose in mg.	Coeff. for log dose (X)	Proportion of cures	Empirical probit	Expected probit, Y	Corrected probit, Y_c	Weight $n_p w_p$	$n_p w_p Y_c$	$n_p w_p Y_c^2$
	20	−1	5/20	4·326	4·27	4·328	10·46	45·27088	195·93237
(S)	40	0	10/20	5·000	5·12	4·999	12·66	63·28734	316·37341
	80	1	17/20	6·036	5·97	6·034	8·97	54·12498	326·59013
			Totals	15·362	—	—	32·09	162·68320	838·89591
	40	−1	2/20	3·718	3·62	3·727	6·18	23·03286	85·84347
(U)	80	0	6/20	4·476	4·68	4·483	12·26	54·96158	246·39276
	160	1	16/20	5·842	5·74	5·837	10·40	60·70480	354·33392
			Totals	14·036	—	—	28·84	138·69924	686·57015

Two yeast extracts are compared in the Table, one of which, the more potent, we will call the standard, and the other the unknown. Twenty birds were given each level of dosage and the doses were equally spaced on a logarithmic scale for both preparations. Empirical probits were summed and the provisional dose-response line computed as described above. The provisional values of b_s and b_u were 0·85 and 1·06 respectively. It was not expected that a second fitting would be necessary, since the deviations of the points from the two provisional lines were small, and that is why the expected probits were determined with some accuracy. The

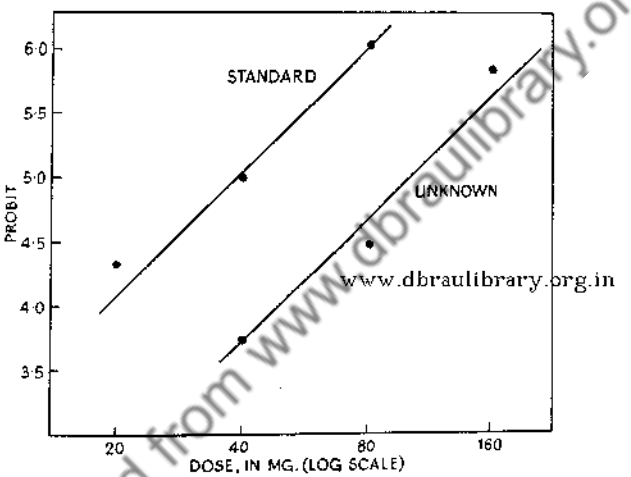


FIG. 16.1. Comparison of the yeast extracts in Table 16.1. A common slope is fitted to both sets of data.

corrected probits differed little from the empirical probits, confirming the goodness of fit of the provisional lines. This example might perfectly well have been computed by method (*b*) above. Method (*a*) is used for illustrative purposes.

We use the coefficients for the logarithm of the dose to simplify the calculations. These are the polynomial coefficients we have already met. The weight factor, w, is multiplied by 20 in all cases to give the weights, weight factors also having been determined by interpolation to the second place of decimals. It is not necessary to tabulate more than two further columns. These are $n_p w_p Y_c$ and $n_p w_p Y_c^2$. Values for $n_p w_p \bar{X}_p$, $n_p w_p \bar{X}_p^2$ and $n_p w_p \bar{X}_p \bar{Y}_c$ are determined by inspection, since the values of X and X^2 are

always plus or minus unity. We then calculate the following quantities:

$$\bar{X}_s = -0{\cdot}04643191 \qquad \bar{X}_u = 0{\cdot}14632455$$
$$\bar{Y}_s = 5{\cdot}0695918 \qquad \bar{Y}_u = 4{\cdot}8092663$$
$$S_s n_p w_p \bar{x}_p^2 = 19{\cdot}43 - 0{\cdot}06918 = 19{\cdot}36082$$
$$S_u n_p w_p \bar{x}_p^2 = 16{\cdot}58 - 0{\cdot}61749 = 15{\cdot}96251$$
$$S_s n_p w_p \bar{x}_p y_c = 8{\cdot}85410 + 7{\cdot}55369 = 16{\cdot}40779$$
$$S_u n_p w_p \bar{x}_p y_c = 37{\cdot}67194 - 20{\cdot}29510 = 17{\cdot}37684$$
$$S_s n_p w_p y_c^2 = 838{\cdot}89591 - 824{\cdot}73742 = 14{\cdot}15849$$
$$S_u n_p w_p y_c^2 = 686{\cdot}57015 - 667{\cdot}04158 = 19{\cdot}52857$$
$$b_s = 0{\cdot}8474739$$
$$b_u = 1{\cdot}0886032$$
$$\chi_s^2 = 14{\cdot}158 - 13{\cdot}905 = 0{\cdot}253$$
$$\chi_u^2 = 19{\cdot}529 - 18{\cdot}916 = 0{\cdot}613$$
$$\chi_b^2 = \frac{(1{\cdot}0886 - 0{\cdot}8475)^2}{\frac{1}{19{\cdot}36} + \frac{1}{15{\cdot}96}} = 0{\cdot}509$$
$$b_c = 0{\cdot}95644$$
$$Vb_c = \frac{1}{19{\cdot}36 + 15{\cdot}96} = 0{\cdot}028313$$
$$s_{b_c} = 0{\cdot}1683$$

Since all values of χ^2 tested indicate no significant departures from homogeneity of responses to samples and slopes, we calculate the combined slope and its error. We note that:

$$\frac{b_c}{s_{b_c}} = \frac{0{\cdot}9564}{0{\cdot}1683} = 5{\cdot}68$$

The provisional values for the two slopes were 0·850 and 1·062. These agreed quite well with the first approximations of 0·847 and 1·089, confirming our belief that one approximation will be all that is necessary. The approximate formula for the errors of M will have to be replaced by a calculation of the exact fiducial limits in any critical examination of the results.

The log ratio of potency, $M = -0{\cdot}19276 - \frac{0{\cdot}26032}{0{\cdot}95644} = -0{\cdot}46494$

Its variance,

$$VM = \frac{1}{0{\cdot}91478}\left(\frac{1}{32{\cdot}09} + \frac{1}{28{\cdot}84} + \frac{(0{\cdot}26032)^2 \times 0{\cdot}028313}{0{\cdot}91478}\right) = 0{\cdot}12725$$
$$s_M = 0{\cdot}2725$$

The value:

$$M \pm s_M = -0{\cdot}4649 \pm 0{\cdot}2725$$

is in logarithms to base 2 and each dose on the unknown is double the corresponding dose on the standard in mgm. Hence, to convert the determination of potency and its standard error to common logarithms, we first multiply M by $\log_{10}2$ and then subtract $\log_{10}2$ from M, and multiply s_M by $\log_{10}2$:

$$\begin{aligned} M \pm s_M &= (-0{\cdot}46494 \times 0{\cdot}30103 - 0{\cdot}30103) \pm 0{\cdot}2725 \times 0{\cdot}30103 \\ &= -0{\cdot}4410 \pm 0{\cdot}0820 \\ &= \bar{1}{\cdot}5590 \pm 0{\cdot}0820 \end{aligned}$$

The approximate limits for $P=0{\cdot}05$ are thus $-1{\cdot}5590 \pm 1{\cdot}960 \times 0{\cdot}0820$, which, taking antilogs, gives antilog $M=0{\cdot}3622$, range 0·2501 to 0·5246. The approximate limits for this level of significance thus indicate that we have established the relative potency with an accuracy of between 69·1% and 144·8% of its most probable value.

A recalculation of these figures using expected probits and weighting coefficients derived from the combined slope, 0·9564, gives practically identical results:

$$\begin{aligned} b_s &= 0{\cdot}8419 \\ b_u &= 1{\cdot}0822 \\ b_c &= 0{\cdot}9594 \\ M &= \bar{1}{\cdot}5587 \pm 0{\cdot}0818 \end{aligned}$$

It should be noted, however, that such a rapid approach to the best fitting line and an accurate estimate of M and its standard error are often not experienced. It may be necessary to make a number of successive approximations before the changes in these values are negligible. The need for a further approximation is always indicated if the difference between successive values of b_c is more than a small fraction of the standard error of b_c, say one-tenth. In this instance, the provisional b_c was $\frac{1}{2}(0{\cdot}85+1{\cdot}06)=0{\cdot}955$, and the first approximation was 0·956, demonstrating no need for further computation of the slope.

16.4. Fiducial limits of error

Fiducial limits of error are calculated for quantal assays in an

analogous way to that explained in Chapter 11.4, with the following modifications:

$$A=\frac{1}{S_s n_p w_p}+\frac{1}{S_u n_p w_p}$$

$$Vb_c=\frac{1}{S_s n_p w_p \bar{x}_p^2+S_u n_p w_p \bar{x}_p^2}$$

In the present example, since $\frac{b}{s_b}$ is only 5·6, we calculate the exact limits, although we do not anticipate any serious discrepancy at the $P=0{\cdot}05$ level. Using the above formulae:

$$A=\frac{1}{32{\cdot}09}+\frac{1}{28{\cdot}84}$$

$$=0{\cdot}06583$$

$$Vb_c=0{\cdot}028313$$

From 11.4, $$C=\frac{0{\cdot}95644^2}{0{\cdot}95644^2-0{\cdot}028313t^2}$$, where $t=1{\cdot}960$

$$=1{\cdot}1350$$

Substituting in the formula for fiducial limits, they are:

$$-0{\cdot}19276-\frac{0{\cdot}26032\times 1{\cdot}1350}{0{\cdot}95644}$$

$$\pm\frac{1{\cdot}960\sqrt{1{\cdot}1350}}{0{\cdot}95644}\left(0{\cdot}06583+\frac{1{\cdot}1350\times 0{\cdot}028313\times 0{\cdot}26032^2}{0{\cdot}95644^2}\right)^{\frac{1}{2}}$$

$$=-0{\cdot}50168\pm 0{\cdot}57000$$

Hence the log limits are $-1{\cdot}07168$ to $0{\cdot}06832$. We convert these limits to common logarithms as before, with due allowance for the difference in dosage scales, whence they are found to be:

$(-1{\cdot}07168\times 0{\cdot}30103-0{\cdot}30103)$ and $(0{\cdot}06832\times 0{\cdot}30103-0{\cdot}30103)$ or $\bar{1}{\cdot}3764$ and $\bar{1}{\cdot}7195$

The corresponding potency limits are: 0·2378 and 0·5242. These fiducial limits range from 65·7% to 144·8% of the mean, and thus differ little from the approximate limits. Note that only the lower limit has in fact shifted.

CHAPTER 17

ASSAYS BASED ON REACTION TIMES

17.1. The use of time as the dependent variable

When either the duration of a response which would otherwise have to be scored as a quantal reaction or the time taken for a response to occur can be recorded, each observation supplies more information than if it had been scored quantally. The speed with which individuals react to the dose of a drug sometimes varies with the dose or with a function of the dose. Alternatively, the time for which they continue reacting may also be related to the dose.

In a simple instance, time may thus take the place of a graded response and the statistical treatment of the results resembles that of assays based on graded responses. An illustration of this method is the use of adrenalectomised drakes in estimating the potency of extracts of the adrenal gland. Drakes are adrenalectomised and then receive injections of the extracts for some arbitrary period, such as one day, and their subsequent period of survival is recorded. The survival time of uninjected adrenalectomised birds averages about 8 to 10 hours, but on the injection of cortical extracts it may be prolonged for several days. In this particular test the time of survival has been found to be linearly related to the logarithm of the dose, and the computational procedure therefore follows exactly the same lines as with any other graded response in which Y is the survival time in any convenient units, such as hours.

Often a linear relationship between the duration of response and the logarithm of the dose is not found, and for various other types of assay the logarithm of the survival time has been found to be linearly related to the logarithm of the dose. The appropriate relationship for any particular test has, therefore, to be determined empirically. In the assay of the vitamin B_1 employing the duration of cure of polyneuritis produced by vitamin depletion in the rat, different authors have related the time in days or the logarithm of the time in days to the dose. The logarithm of the time seems to be related with more frequent success to the log dose than the time itself.

17.2. Reaction time at a given response level

Where it is inconvenient or impracticable to score the individual reaction times in full, it has been a frequent practice to determine the reaction time in groups of test objects for a given level of response, such as 50 % mortality. This may be done by repeated observations at fixed intervals on a single group, given a standard dose of the substance being tested, or by determining at a constant time or times after treatment the percentage response in groups receiving various separate doses. In the assay of tetanus anti-toxin at the State Serum Institute, Copenhagen, the time at which 50 % of a group had died was chosen as the response. This time was determined for each group of animals by observing the number of survivors daily for six days and then determining t_{50}, the time at which 50 % had died, according to the formula:

$$t_{50} = t_{100} + \frac{N}{2} \times \frac{S(t - t_{100})}{SN - Nt}$$

where N equals the number of animals injected, t_{100} equals the last day on which 100 % survived, $t - t_{100}$ equals any day reckoned from t_{100}, N_t equals the survivors on the tth day and $N - N_t$ equals the total number of deaths up to the tth day. The curves obtained were not fitted mathematically and thus no critical examination was made of the adequacy of this method of scoring results.

In general, a more accurate way of dealing with such a problem is to determine a series of curves relating mortality to time after treatment at various dose levels or a series of curves relating mortality to dose at various times after treatment. In such investigations it has, for instance, been found that the reaction time for a given percentage response is linearly related to the logarithm of the dose. Various methods of relating, first, one of the three variables involved to the second by means of regression lines and then relating some arbitrary level of response or time of reaction computed from these lines to the third variable have been tried, the commonest of which has been to relate the time taken to achieve a certain level of response to the log dose. In the process of determining such relationships, a number of mathematical assumptions and semi-arbitrary corrections have frequently been found to be necessary and the amount of time and labour involved in the calculations must often have been considerable. None of these methods is really satisfactory from a mathematical point of view. There is, on the other hand, a more elegant and satisfactory method

of dealing with the problem which will generalise easily, however many factors are involved.

17.3. Partial regression

Before passing to the more general solution of this type of problem, we must consider the concept of partial regression. We shall not deal with the completely general case, but confine our attention to two independent variates.

When the response depends on two factors, such as dose and duration of treatment, we may write the relationship between them according to the formula:

$$Y=a+b_1X_1+b_2X_2$$

where b_1 and b_2 are called *partial regression coefficients* and the equation is the partial regression equation. This formula assumes that the two factors influencing any particular response are linearly related to it. By analogy with the equation involving a single regression coefficient, we find the two equations:

$$b_1Swx_1^2+b_2Swx_1x_2=Swx_1y$$
$$b_2Swx_2^2+b_1Swx_1x_2=Swx_2y$$

where w is the total weight of an observation ($w=n_pw_p$), Swx_1^2 is the sum of the products of each weight factor and the square of the deviation of X_1 from $\bar{X}_1$, and the other sums of squares and products have analogous meanings, as in our usual notation, but omitting subscripts in a general equation of this type. These equations can be solved directly to give the values of b_1 and b_2, but it is more useful to apply a method known as the *inverse matrix*, since by using this method the variances may be more readily obtained.

$$\text{If } b_1=c_{11}Swx_1y+c_{12}Swx_2y$$
$$\text{and } b_2=c_{12}Swx_1y+c_{22}Swx_2y$$

we determine the values of c_{11} and c_{12} from the following relationships:

$$c_{11}Swx_1^2+c_{12}Swx_1x_2=1$$
$$c_{11}Swx_1x_2+c_{12}Swx_2^2=0$$

We then determine c_{12} and c_{22} from:

$$c_{12}Swx_1^2+c_{22}Swx_1x_2=0$$
$$c_{12}Swx_1x_2+c_{22}Swx_2^2=1$$

A check on the calculations is provided by the duplicate determination of c_{12}. By the use of c multipliers we may test the significance of any linear function of two or more regression coefficients by

calculating its standard error, using the t Table (or the normal deviation in the case of quantal responses).

17.4. Application of partial regression coefficients to dose-response data with two dosage factors

We are now in a position to examine the application of this method to dose-response data. This application involves a generalisation of the analysis in which one independent variate is related to the response. We first determine empirical probits by the following methods:

1. If the experiment has been arranged so that a series of doses is tested at each of several different times, we may draw provisional lines, one to each of the dose-response relationships, for each separate time interval, using the probit against log dose, but these lines should be parallel to one another and at distances apart proportional to the differences in log time.
2. If the experiment has involved a series of tests each giving the response to a constant dose at variable times, we plot a number of provisional probit-log time lines parallel to one another and at distances apart proportional to the log dose.
3. If a series of doses and series of times are used in all or nearly all of the possible combinations, the probits should be plotted against the sum of log dose and log time (X_1+X_2). Points with constant X_2 are fitted by one set of parallel lines and points with constant X_1 are fitted by a second set of parallel lines, and a plane representing the three-dimensional relationship of probits to log dose and log time is obtained and provisional probits are read from the two intersecting sets of parallel lines drawn by eye, or provisionally calculated, if the logarithmic intervals make this easy.

Care over drawing the provisional lines should ensure that a satisfactory fitting is obtained with the minimum number of stages in computation. Corrected probits and weights are derived from the expected probits and the percentage response, exactly as when only one factor is involved. If the equation calculated from the first approximation differs substantially from that appropriate to the provisional lines, which should have slopes near to b_1 and b_2, it will be necessary to calculate a further approximation, and so on.

For the standard and unknown in an assay we shall then have two probit planes instead of lines as when one factor is involved; each plane is a *regression plane*. These regression planes can be tested

for homogeneity and parallelism just as with two straight lines by means of χ^2 or the analysis of variance. The way to do this will be explained in the practical example which follows.

17.5. An analysis with probit planes

Data obtained at Rothamsted Experimental Station on the toxicity of a pyrethrum oil spray on *Tribolium castaneum* have been used by D. J. Finney (*Ann. Appl. Biol.*, **30**, 71, 1943) as an example of the arithmetical procedure of fitting probit planes. Finney's analysis is used here as an illustration of this procedure, which he himself elaborated. This particular test compares the percentage kills of *T. castaneum* when the poison was sprayed directly onto the insects and when the insects were placed on a disc which had previously been sprayed so as to cover it with a film of poison. Four different concentrations of poison in oil were sprayed, so as to give three different final concentrations of deposit (1) as a direct spray and (2) as a film, so that the two variables, X_1 and X_2, are concentration of spray in mg./ml. and the strength of the deposit in mg./sq. cm., for the direct spray and film techniques respectively. This example does not contain observations of a time factor, but the technique of analysis is exactly the same as if, for instance, the four concentrations had been tested for three different periods of time, the insects being exposed to the spray during each of these three periods. No practical test using time as one of the variables appears to have been reported in which the results lend themselves so well to analysis as in this particular experiment.

Tests made with the oil alone, in which the poison is normally sprayed, killed 3·9% of insects, and thus adjustment had to be made to each observed mortality, according to the formula:

$$p = \frac{100(p_o - p_c)}{100 - p_c}$$

where p_c and p_o are respectively the observed control and experimental percentage kills and p is the adjusted kill. The results of the test after correction for the mortality amongst the controls are shown in Table 17.1.*

* This book was in press when Mr. Finney's volume *Probit Analysis* (Cambridge University Press, 1947) appeared. Although references to this work have been inserted it has not been possible to do full justice to the many refinements it discusses. Among these is a recalculation of the *T. castaneum* data with weighting factors corrected for reactions in the controls, which even at a 4% level of control mortality affects the estimate of Δ and its variance to a surprising degree. The new value for Δ is 0·203 and its variance is 0·0216, to be compared with 0·169 and 0·0191 on page 179. Tables of weighting factors for percentage of control reactions from 0 40% are given in Finney's book.

TABLE 17.1

PERCENTAGE KILLS OF *T. castaneum* BY A PYRETHRUM OIL SPRAY, ADJUSTED FOR 3·9% MORTALITY AMONGST THE CONTROLS

(Numbers of insects shown in brackets)

(From Finney, *Ann. Appl. Biol.*, **30**, 71, 1943)

Conc.	Deposit (mg./sq. cm.)					
	Spray			Film		
mg./c.c.	0·29	0·57	1·08	0·29	0·57	1·08
0·5	0·0 (27)	10·3 (29)	16·8 (30)	6·7 (29)	11·4 (27)	25·7 (28)
1·0	49·8 (29)	64·1 (29)	61·0 (24)	30·7 (30)	48·0 (28)	59·1 (28)
2·0	89·6 (30)	96·1 (27)	100·0 (31)	82·1 (29)	96·3 (28)	92·6 (28)
4·0	100·0 (28)	100·0 (30)	100·0 (19)	100·0 (29)	100·0 (29)	100·0 (17)

Expected probits were determined by the method outlined in the preceding section from a figure similar to that shown in Figure 17.1,

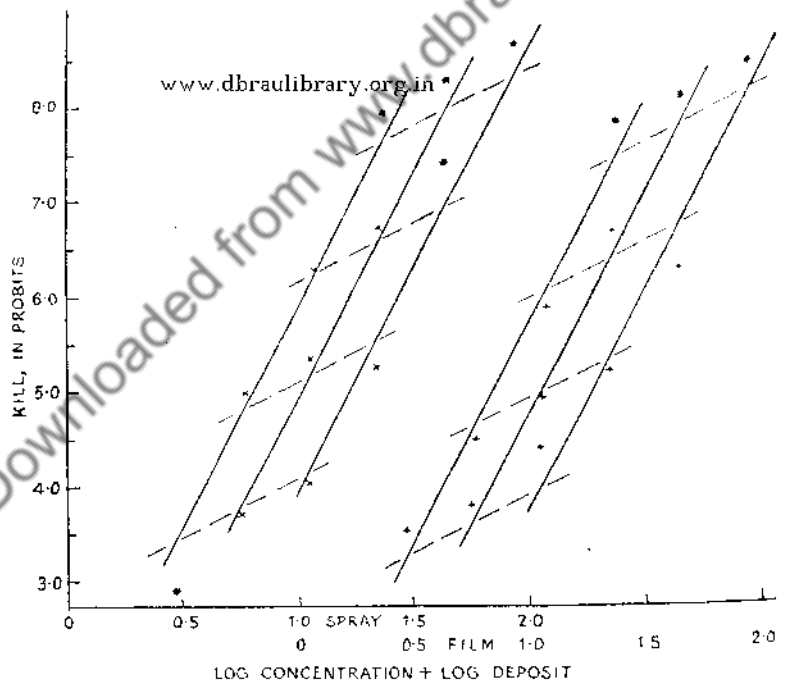


FIG. 17.1 Toxicity of pyrethrum oil spray to *T. castaneum*. × direct spray technique; ÷ film technique; * 0% or 100% kill; ——— effect of change in concentration for fixed deposit; - - - - effect of change in deposit for fixed concentration. (From Finney, *Ann. Appl. Biol.*, **30**, 71, 1943.)

TABLE 17.2

COMPUTATIONS FOR FITTING OF PROBIT PLANE FOR TOXICITY OF PYRETHRUM OIL SPRAY TO *T. castaneum* (DIRECT SPRAY TECHNIQUE)

(From Finney, *Ann. Appl. Biol.*, **30**, 71, 1943)

X_1	X_2	n_p	% kill	Empirical probit	Expected probit	w $(=n_p w_p)$	Y_c	wX_1	wX_2	wY_c	E
0	0·47	27	0	$-\infty$	3·4	6·4	2·91	0		18·624	3·44
0·30	0·47	29	49·8	5·00	4·8	18·2	5·00	5·46		91·000	4·84
0·60	0·47	30	89·6	6·26	6·3	10·1	6·26	6·06		63·226	6·25
0·90	0·47	28	100·0	∞	7·6	1·1	7·94	0·99		8·734	7·65
						35·8		12·51	16·826	181·584	
0	0·75	29	10·3	3·74	3·8	10·7	3·74	0		40·018	3·79
0·30	0·75	29	64·1	5·36	5·2	18·2	5·36	5·46		97·552	5·19
0·60	0·75	27	96·1	6·76	6·6	6·4	6·74	3·84		43·136	6·60
0·90	0·75	30	100·0	∞	8·0	0·4	8·30	0·36		3·320	8·00
						35·7		9·66	26·775	184·026	
0	1·04	30	16·8	4·04	4·1	14·1	4·04	0		56·964	4·15
0·30	1·04	24	61·0	5·28	5·5	13·9	5·27	4·17		73·253	5·55
0·60	1·04	31	100·0	∞	7·0	4·1	7·42	2·46		30·422	6·96
0·90	1·04	19	100·0	∞	8·4	0·1	8·67	0·09		0·867	8·36
						32·2		6·72	33·488	161·506	
						103·7		28·89	77·089	527·116	

which in fact gives the final result of fitting. Weighting coefficients and corrected probits were calculated exactly as described in Chapter 14 and the first stage of the calculations is shown in Table 17.2 for the fitting of the probit plane for the direct spray technique. In this Table, $\log(2 \times X_1)$ and $\log(10 \times X_2)$ have been used for convenience in working. From the data in the Table and from further columns giving SwX_1^2, SwX_1X_2, SX_2^2, SwX_1Y_c, SwX_2Y_c and SwY_c^2, the values of the sums of squares and products of deviations were calculated as listed in Table 17.3. We continue to omit the rather ponderous subscripts—thus SwX_1X_2 would, in our usual notation, be $Sn_pw_p\bar{X}_{p1}\bar{X}_{p2}$, but the meaning should be clear. All computations of sums of deviations of squares and products for both methods of applying the poison were computed by the usual techniques. The sums of squares and products in Table 17.3 are the totals for all 12 doses of $X_1 + X_2$ given by each technique.

TABLE 17.3

SUMS OF SQUARES AND PRODUCTS FOR EXPERIMENT WITH *T. castaneum*

(From Finney, *Ann. Appl. Biol.*, **30**, 71, 1943)

Description	Spray	Film	Total
Swx_1^2	5·190	5·437	10·627
Swx_1x_2	−1·363	−1·326	−2·689
Swx_2^2	5·510	5·821	11·331
Swx_1y_c	25·390	21·018	46·408
Swx_2y_c	−0·520	2·043	1·523
Swy_c^2	135·23	95·57	230·80

The sums for the two techniques together are then added and are given in the third column of the Table, since the provisional probit lines suggested that the two corresponding lines for the two techniques were parallel and thus that two parallel probit planes could most probably be fitted to the data. From this column of totals the regression coefficients b_1 and b_2 were calculated by means of the inverse matrix, and it was found that:

$$c_{11} = 0{\cdot}10011$$
$$c_{12} = 0{\cdot}02376$$
$$c_{22} = 0{\cdot}09389$$

whence:

$$b_1 = 46{\cdot}408c_{11} + 1{\cdot}523c_{12} = 4{\cdot}682$$
$$b_2 = 46{\cdot}408c_{12} + 1{\cdot}523c_{22} = 1{\cdot}246$$

The total sum of squares of the deviations of Y_c, Swy_c^2, was 230·80 and of this total, the amount represented by the fitting of parallel planes to two sets of points—analogous with the sum of squares due to linear regression in a one-factor experiment—is:

$$b_1 Swx_1 y_c + b_2 Swx_2 y_c = 219{\cdot}18$$

The separate values of b_1 and b_2 for the spray and film methods respectively were also calculated from the separate totals for the spray and film methods and found to be 5·205 and 1·193 for the spray technique and 4·184 and 1·304 for the film technique. The sum of squares attributable to the fitting of two separate planes was then calculated exactly as above, but separately for each technique, and found to be 131·53 for the film and 90·60 for the spray, totalling 222·13. The difference between this sum of squares and 219·18, the sum of squares accountable by the fitting of parallel planes, is 2·95, and this is a measure of departure from parallelism. The residual sum of squares is 230·80 − 222·13 = 8·67, and this is a measure of the heterogeneity of the points about the parallel planes which have been fitted to them. The total sum of squares was thus split up as in Table 17.4, in which the degrees of freedom were determined by the following considerations.

TABLE 17.4

ANALYSIS OF χ^2 FOR THE *T. castaneum* ASSAY

	Degrees of freedom	Sum of squares	Mean square
Regression plane	2	219·18	
Parallelism of planes	2	2·95	
Heterogeneity	18	8·67	0·48
Total	22	230·80	

The fitting of parallel planes requires two constants and this removes two degrees of freedom.

The fitting of two distinct planes requires the calculation of four constants and removes four degrees of freedom, two in addition to the degrees already used in the computation of parallel planes.

As there were 11 degrees of freedom available in each set of results, there remain 18 degrees of freedom associated with the residual sum of squares of 8·67. The 23rd degree of freedom, representing differences between substances (techniques), does not enter into this Table. Remember that Table 17.4 is not, strictly

speaking, an analysis of variance, but must be tested by χ^2. The residual sum of squares, that associated with possible heterogeneity, is tested by χ^2 with 18 degrees of freedom; for the total data the value of $\chi^2_{[18]}$ would have to exceed 28·87 to indicate heterogeneity at the 5% level. There is clearly no suggestion of any heterogeneity in these results. In consequence the significance of the departure of the two planes from parallelism is tested with two degrees of freedom by χ^2, and found to be well below the 5% level.

If the χ^2 test had shown significant departures from homogeneity, but departures of such a kind that a different relationship from that tested was not apparent from the data, so that the assumption of parallel planes would not be rejected, the test of parallelism would be made by an F-test, in which the mean square for parallelism would be compared with the mean square for heterogeneity. Then, to allow for the heterogeneity in calculations of error, all the variances in the subsequent analysis would have to be multiplied by the mean square for heterogeneity, as in the case where the same phenomenon is seen in dealing with ordinary probit-log dose-response data.

It may be noted that the possibility that χ^2 might indicate heterogeneity if calculated by a more exact method (cf. Chapter 15) was not examined. In view of the small value found for the $\chi^2_{[18]}$ above, it is reasonable to neglect this possibility, as this value would have to be associated with as few as three degrees of freedom before significance at the 5% level could be attributed to it.

The dose-response relationship, i.e. the equations of the probit planes, may be written:

$$E=\bar{Y}_c+b_1(X_1-\bar{X}_1)+b_2(X_2-\bar{X}_2)$$

each $\bar{X}$ and $\bar{Y}$ being taken separately for each technique, but the common values of b_1 and b_2 are used. The equations found were:

$$E_s=2{\cdot}853+4{\cdot}682X_1+1{\cdot}246X_2$$

and $$E_f=2{\cdot}684+4{\cdot}682X_1+1{\cdot}246X_2$$

where E_s and E_f are the estimates for spray and film technique respectively. The variance of the mean probit for each technique is given by:

$$V\bar{Y}=\frac{1}{Sw}, \text{ (where } w=n_p w_p)$$

The variance of b_1 is c_{11}, the variance of b_2 is c_{22} and the covariance of b_1 and b_2 is c_{12}. (When dealing with a graded response the

corresponding variances and covariance are $c_{11}Ve$, $c_{12}Ve$ and $c_{12}Ve$.) From these the variance of the predicted probit E for any dosage is:

$$VE=V\bar{Y}+c_{11}(X_1-\bar{X}_1)^2+2c_{12}(X_1-\bar{X}_1)(X_2-\bar{X}_2)+c_{22}(X_2-\bar{X}_2)^2$$

The last column of Table 17.2 gives the values of E for the direct spray technique; these are little different from the expected probits, showing no necessity to repeat the operation of fitting.

17.6. Relative potency

In experiments relating one dosage factor to response, it is usual to compare the relative potencies of two substances by the antilogarithm of the horizontal distance between the two lines. When probit planes have been fitted, similar computations for each of the two variables separately, holding the second constant, may be made. In a test involving both a reaction time and dosage, the ratio of times of exposure and of doses eliciting the same percentage reactions measure different aspects of relative potency. If the probit planes are parallel, these relative potencies are independent of the level of response chosen for making the comparison. Thus, although we may choose to return to comparisons in terms of one or other factor alone, the fitting of probit planes will nevertheless have enabled us to compute relative potencies with greater precision and to make comparisons by other methods if we afterwards feel them to be necessary.

Finney suggests that a more useful comparison is given by the *mean probit difference*, denoted by Δ. The mean probit difference is the difference between the predicted probits for the two substances at given dosages. When the regression planes are parallel, Δ is independent of dose level and is the vertical distance between the two planes. In terms of the standard and unknown:

$$\Delta=\bar{Y}_s-\bar{Y}_u-b_1(\bar{X}_{1s}-\bar{X}_{1u})-b_2(\bar{X}_{2s}-\bar{X}_{2u})$$

with a variance:

$$V\Delta=\frac{1}{Sw_s}+\frac{1}{Sw_u}+c_{11}(\bar{X}_{1s}-\bar{X}_{1u})^2+2c_{12}(\bar{X}_{1s}-\bar{X}_{1u})(\bar{X}_{2s}-\bar{X}_{2u})+c_{22}(\bar{X}_{2s}-\bar{X}_{2u})^2$$

In the example just considered $\Delta=0{\cdot}169$ and $V\Delta=0{\cdot}0191$.

17.7. Interaction

The equation:

$$Y=a+b_1X_1+b_2X_2+b_{12}X_1X_2$$

is a more general equation, which, however, may still be reduced to a linear form when one of the Xs is held constant. If such an equation describes a dose-effect relationship, it means that the effects of the separate factors are not independent and additive, but that although Y is linearly dependent on either dose factor separately, the slope of the regression of Y on X_1 increases as X_2 increases if b_{12} is positive and decreases if it is negative. The regression coefficient b_{12} represents the interaction of the two factors. The fitting of surfaces of this form can be done by simple extension of the methods used above, but the possibilities offered by the equation do not yet seem to have been realised in practice. In one case in the literature examined by Finney, where the fit given by probit planes was significantly less good than it should have been, calculations including interaction did not improve matters.

CHAPTER 18

GROUPS OF TESTS

18.1. Groups of simultaneous tests

It is often desirable to determine the potency of several substances at the same time. The object of the test may be to save time and labour by using one determination of responses to the standard for comparison with several different unknowns, or to compare the relative potencies of several substances with one another in the absence of any particular standard.

If several restrictions are to be included in the design of the test, as in a test based on the Latin square, we must choose a design in which all dosage levels of all substances contribute equally to the totals of responses in rows and columns or to totals for any other restrictions in design that may have been imposed. However, it is usually difficult to cope with many substances at once in such a design, as the number of cells required in the square mounts rapidly with the number of substances to be tested. This leads to difficulties in planning the assay, particularly if litter-mates or similar groups of limited numbers of test objects are to be used.

Methods of surmounting some of these difficulties will have been suggested by the designs discussed in Chapter 13, but it must be admitted that, in practice, the simultaneous testing of several substances usually necessitates simplification in the design of the assay or the use of balanced incomplete blocks. Thus, the advantage of testing several substances together may be offset by the loss of a certain degree of precision.

The results of a group of tests conducted simultaneously can be pooled to give estimates of the slope of the dose-response line and of the error variance, as long as there is no reason to believe that any one substance is evoking an atypical response. Thus, the slope of the line for each substance may need to be examined separately if the results suggest that there may be significant differences between slopes, or the responses to one substance may best be compared separately with those to the standard if the variance of these responses is significantly greater than that for the other substances.

Such rejections must only be made on good grounds, for in the absence of significant differences between slopes or variances, the best estimate of the slope and error variance for the assay is that derived from all substances together. Methods for testing differences between variances have been described in Chapter 9.6, and methods for testing differences between slopes are discussed below.

18.2. Comparison in a group of tests

The following example shows how a group of substances may be compared simultaneously with a standard. By the "cylinder plate" method of assaying the potency of penicillin a warm agar medium is seeded with *Staphylococcus aureus* and poured into a shallow container and allowed to set around a series of small upright cylinders, into each of which a quantity of solution of penicillin is placed. The penicillin diffuses into the surrounding medium and inhibits the growth of *S. aureus* over a circular area, the diameter of which may be used as the response to the drug. This diameter increases with increasing amounts of penicillin and is linearly related to the logarithm of the dose over a wide range. In the present test, three preparations of penicillin are compared with a standard, each at three dose levels, and three cylinders were used at each level, a total of 36 cylinders, all on one agar spread. The diameter of each area of growth inhibition is given in thousandths of an inch in Table 18.1.

The analysis of variance takes the usual form. There are 24 degrees of freedom for the estimation of error—two within each cell of the Table. The actual cylinders were distributed over the agar surface, not grouped as in the Table. This point will be referred to below when discussing the results. Differences between substances account for a further three degrees of freedom, leaving eight degrees of freedom which will be accounted for in the examination of the slope and departures from linearity and parallelism. These eight degrees of freedom may be conveniently grouped for preliminary analysis as follows:

Source of variation	Degrees of freedom
1. Linear regression	1
2. Combined curvature of dose-response lines	1
3. Departures from parallelism and opposed curvature of separate dose-response lines	6
Total	8

TABLE 18.1

AN ASSAY OF THREE SAMPLES OF PENICILLIN BY THE CYLINDER PLATE METHOD

Sample	Dose in ml.			Totals (T_s)
	0·8	1·0	1·25	
Standard	607	673	740	
	577	615	645	
	605	643	700	
Totals (T_p)	1789	1931	2085	5805
U_1	614	661	742	
	582	615	652	
	590	630	689	
Totals (T_p)	1786	1906	2083	5775
U_2	608	652	758	
	580	611	633	
	605	630	684	
Totals (T_p)	1793	1893	2075	5761
U_3	606	661	723	
	568	596	637	
	562	584	659	
Totals (T_p)	1736	1841	2019	5596
Totals (T_d)	7104	7571	8262	$22937 = T$

Items 1 and 2 above are calculated from the three totals (or means) of all responses to each of the three dose levels and thus refer to the combined slope and combined curvature of all four curves. Item 3 is the difference between the sum of squares for the 12 cell totals of the Table and items 1 plus 2 plus the sum of squares for differences between substances (it is thus the *dose/substance interaction*—see Chapter 12.4). The analysis of variance thus takes the form shown in Table 18.2.

It will be seen that the sum of squares for dose/substance interaction is much less than would be expected by chance. This does not imply, as it might in some circumstances, that the results have been "cooked" to make them more consistent and that we have uncovered this charlatanism by statistical analysis, but almost certainly derives from a peculiar circumstance of this particular assay. The three cylinders containing the equal doses of any one preparation were not distributed at random over the medium, but were distributed systematically, so that they sampled the inequalities

TABLE 18.2

ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 18.1

Source of variation	Formula	Degrees of freedom	Sum of squares	Mean square	F	P
Between doses	$Sn_d\bar{y}_d^2$	(2)				
A. Linear regression	$\frac{(Sn_d\bar{x}_d\bar{y}_d)^2}{S\bar{x}_d^2}$	1	55873·5	55873·5	44·2	<0·001
B. Combined curvature	$Sn_d\bar{y}_d^2-A$	1	696·9	696·9	0·55	>0·05
C. Between substances	$Sn_s\bar{y}_s^2$	3	2943·9	981·3	0·78	>0·05
D. Dose/substance interaction	$Sn_p\bar{y}_p^2-(A+B+C)$	6	193·7	32·3	0·025	<0·001
Error	$Sy^2-Sn_p\bar{y}_p^2$	24	30332·4	1263·8	—	—
Total	Sy^2	35	90040·4	—	—	—

of the medium more effectively than would usually have occurred by chance. However, since they were not arranged in a design, such as a Latin square, in which a sum of squares for such items as row or column variation could be eliminated from that attributable to error, this balanced design has effected an increase in real precision accompanied, paradoxically, by a decrease in apparent precision. Such variation as occurred between parts of the surface of the medium is represented excessively in the sum of squares for error—which is no longer properly named "random sampling," for the sampling was not at random. It is of interest that the present author was unaware of the physical lay-out of this assay when analysing it, and only after the above finding suspected and established that a non-random arrangement had been employed and that the sensitivity of the assay is thus impaired.

Assuming an unbiassed estimate of error, however, we can proceed without further analysis and use the combined error sum of squares and the combined estimate of slope for determining the errors of the three assays of substances U_1, U_2 and U_3 against the standard by the usual means, with the advantage of a larger number of degrees of freedom for the estimation of these errors than would have been available had each to be compared separately with the standard. Fiducial limits of error will be narrowed in so far as s_b is reduced by combining estimates of b from all substances.

18.3. Significant heterogeneity in slopes

Had the mean square for dose/substance interaction in Table 18.2 proved significantly greater than that for error, we should conclude that the individual dose-response lines are not sufficiently alike to be treated as samples from the same population of such lines. We must then examine in more detail the contributions to the six degrees of freedom associated with the interaction in order to eliminate the line or lines responsible for the excessive variability. The method has been discussed in Chapter 12, but the analysis in the present instance, where the presence of replication gives 24 degrees of freedom not associated with segregable interactions between restrictions in design, is simpler.

The sum of squares attributable to opposed curvature of the four individual dose-response lines, with which are associated three degrees of freedom, is:

$$\frac{S(T_{p1}+T_{p3}-2T_{p2})^2}{6n_p}-\frac{(T_{d1}+T_{d3}-2T_{d2})^2}{6n_d}=D1$$

where T_{p1} = the total of responses to the lowest dose of each substance,

T_{p2} = the total of responses to the middle dose of each substance,

T_{p3} = the total of responses to the highest dose of each substance.

T_{d1}, etc., are the corresponding totals for all substances together, and n_p and n_d are the numbers of observations in each group. The sum of squares attributable to departures from parallelism, with which are associated the remaining three degrees of freedom, is:

$$\frac{S(T_{p3}-T_{p1})^2}{2n_p}-\frac{(T_{d3}-T_{d1})^2}{2n_d}=D2$$

and D1+D2 should, as a check, add up to the interaction sum of squares, D, in Table 18.2.

Each mean square, when tested against the error mean square in the F-test (with three and 24 degrees of freedom respectively in the present example), may differ significantly from it, and inspection of the individual items calculated in forming the first half of the left-hand side of each of the above equations will show which of them is contributing most to the sum of squares concerned. Note that this method, in which the factorial coefficients of Chapter

10 have been employed, is only applicable with a balanced design, with equal numbers of observations per group and equal spacing of log doses. Without this, full analysis will be extremely tedious or even impossible.

Assays in which covariance analysis has also been employed can be examined in just the same way, using reduced sums of squares as explained in Chapter 8.

18.4. Series of tests made over a period

Another frequent occurrence in assay work is the accumulation of a series of results in which the same or different substances have been compared from time to time by the same test method. We may be interested in combining such tests to give an improved estimate of the slope of the typical dose-response line or of the potency of an unknown.

It is commonly found that, whereas the slope of the dose-response line remains substantially constant, the response to a given dose of the standard varies from time to time. This is not always true, but it is so likely to occur that it would be unwise to assume that time-to-time variation in response can be neglected unless strong evidence to the contrary is forthcoming. Thus, we may often be justified in combining the slopes of various tests made at different times, but rarely justified in combining a series of responses to a standard dose. We can, however, always combine a series of estimates of potency in order to arrive at a mean estimate, which, if the test is reliable, will have a greater precision than any of the individual estimates.

18.5. The combined estimate of error

When a series of tests has been conducted by the same technique and the mean square for error has been determined after the elimination of the same sources of variation in each instance, we may combine the estimates of *Ve* by pooling all residual sums of squares and dividing by the added degrees of freedom with which they are associated.

This combined estimate ($\overline{Ve}$) may be used in computing the errors of combined estimates of the slope and of potency in sections 18.6 and 18.7 and in predicting the mean square for error to be expected in future tests. A combined estimate of the error variance will, of course, automatically be computed when a series of tests can be subjected to an overall analysis, in which time-to-time

differences in response are automatically eliminated in the estimation of potency and its error.

18.6. The combined estimate of the slope

If we have a series of r estimates of b, the slope, the mean of these, $\bar{b}$ may be estimated from the weighted mean of the separate estimates. Each value of b is weighted inversely according to its variance, or alternatively, the weighted mean may be computed from:

$$\bar{b}=\frac{SSn_pw_p\bar{x}_p\bar{y}_p}{SSn_pw_p\bar{x}_p{}^2}$$

and

$$V\bar{b}=\frac{V\bar{e}}{SSn_pw_p\bar{x}_p{}^2}$$

Note that with graded responses w_p will in general be unity, and with quantal responses $V\bar{e}$ will usually be unity.

Differences between individual estimates of the value of the slope may be significant. If the series of tests has been of balanced design, an analysis of variance covering the whole series and segregating sources of variation which can be eliminated from the estimate of error will detect any such differences. If we are pooling a series of estimates from tests which cannot easily be combined in this way, we compare the two estimates of the variance of the mean slope:

(a) $$V\bar{b}=\frac{V\bar{e}}{SSn_pw_p\bar{x}_p{}^2}$$

and (b) $$V\bar{b}=\frac{Sw(b-\bar{b})^2}{Sw(r-1)}, \text{ where } w=\frac{1}{Vb}.$$

If (a) is significantly less than (b), the observed variance is greater than that predicted from our knowledge of the individual determinations of the slope, and thus the slope is not constant over the series of assays.

This comparison may be made with the χ^2 table in both quantal and other cases:

$$\chi^2_{[r-1]}=\frac{SSn_pw_p\bar{x}_p{}^2(b-\bar{b})^2}{V\bar{e}}$$

which is a modified form of

$$\chi^2_{[r-1]}=S\frac{b^2}{Vb}-\bar{b}S\frac{b}{Vb}$$

18.7. The combined estimate of potency

A combined estimate of potency is made by analogous methods. The weighted mean log potency should have a variance $V\bar{M}$ given by:

$$(c)\quad \frac{1}{V\bar{M}}=S\frac{1}{VM}$$

which should not differ significantly from the alternative estimate:

$$(d)\quad VM=\frac{Sw(M-\bar{M})^2}{Sw(r-1)},\ \text{where } w=\frac{1}{VM}.$$

Using χ^2 as before:

$$\chi^2_{[r-1]}=S\frac{M^2}{VM}-\bar{M}S\frac{M}{VM}$$

If the variance of the mean log potency is significantly greater when measured directly from the series of estimates under examination than when computed from the combined data of the individual tests, it means that the method of assaying these substances is unreliable. However, the best estimate of potency available remains the weighted mean, which has a variance given by equation (d).

18.8. The use of combined estimates in future tests

If it has been shown that the slope of a dose-response line does not vary significantly, use may be made of the fact in assessing the results of further assays, as long as the further evidence added by the new assays does not disturb previous conclusions as to the constancy of the slope. Furthermore, for pilot tests of new substances or rapid surveys of the potency of several preparations at once, the mean of previous slopes may be used without always determining the slope anew. It is perfectly permissible under these conditions to make pilot assays with only one dose of an unknown, although it is best always to include two doses of the standard, so that evidence is continually accumulating about the slope of the dose-response line. Any full determination of potency must, however, ensure that the slope of the dose-response line is the same for the unknown as for the standard. If it is for any reason imperative to attempt an estimation of potency in the first test made with a new substance, two doses at least must be employed, unless the exact chemical constitution of the substance is known and only its degree of dilution is in doubt.

18.9. Combined estimate of linearity of regression

If r tests have been made, with r different estimates of the value of b, we may wish to test the validity of the assumption that log dose is linearly related to response. For this purpose we compare $V\bar{e}$, derived from all tests, with the variance attributable to deviations from regression, which is such that:

$$Vr = \frac{S(Sn_pw_p\bar{y}_p^2 - b_tSn_pw_p\bar{y}_p\bar{x}_p)}{S(n_t-2)}$$

where w_p, $\bar{y}_p$, etc., refer to the weight and mean response, etc., to dose group, p, in test t, of which b_t is the slope, based on observations from n_t groups. There will be $S(n_t-2)$ degrees of freedom available for the comparison with $V\bar{e}$ in the F-test. If Vr significantly exceeds $V\bar{e}$, we cannot maintain the hypothesis that the log dose-response relationship is linear, and that Ve adequately measures error. Since this test will only be used when a rather heterogeneous series of assays is combined—otherwise an overall analysis of variance can be used—general inspection of the data must be employed in deciding whether the dose-response line is curved or whether the individual mean responses depart more than is to be expected from a supposedly linear relationship.

In assays employing a quantal response, the above test is most easily performed by adding together the values of χ^2 for heterogeneity; the added values are again tested as χ^2 with $S(n_t-2)$ degrees of freedom. (If anyone is worried by the sudden introduction of χ^2 into assays based upon graded responses in Sections 18.6 onwards, he may be reassured by remembering that, as long as n_2 is large, $\chi^2 = En_t$.)

18.10. Fiducial limits of error

When combining a heterogeneous series of tests, if $\frac{\bar{b}}{s_{\bar{b}}}$ is greater than 8, fiducial limits of error need not be calculated, as $\bar{M} \pm ts_{\bar{M}}$ will be sufficiently accurate. The calculation of fiducial limits when this is not the case, or when b varies from test to test, has yet to be described, except in a few particular instances. It is always possible to calculate such limits when the whole series of tests is of balanced design, and the problem is only likely to arise in acute form when a hotch-potch of tests has to be handled, and the estimate of potency derived from the whole series is clearly subject to large errors. The obvious answer at the moment is not to arrange one's own tests so that this situation is likely to occur.

CHAPTER 19

CHOOSING AND MEASURING THE RESPONSE

19.1. The choice of responses

Where there is a choice in the response or method of measuring the response in a biological assay, we naturally wish to find the response which gives the most accurate assay and is not at the same time too laborious to measure or compute. In general, a simple measure of response is as likely to give results as accurate as more complex and involved measurements.

A quantitatively measurable response gives more information per animal than a quantal reaction, and the latter is thus rarely to be preferred to the former. Where responses can be arranged in order of magnitude or in grades, but not given a numerical score other than some purely arbitrary value, it is possible to use the method of Fisher (*Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh), by which numerical values which will best differentiate between doses of a drug are assigned on the assumption that the individual effective doses are normally distributed, but the arithmetic is rather forbidding.

The choice of criteria will also depend on a number of other factors—the mean squares for error, linearity of regression, the steepness of the dose-response lines, the amount of work involved in alternative types of measurement, and the statistical method by which the results will, or can, be treated. The latter may be of particular importance, as in some tests so designed that an analysis of variance and covariance is possible, it may not matter, as far as the accuracy of the test is concerned, which criterion is chosen, since the efficiency of the statistical procedure may be such as to yield equally accurate estimates of potency in all cases. Where such an analysis must, by the nature of the data, be incomplete, it is important to know whether one method is likely to give greater accuracy than another, and how this accuracy is affected by such corrections as adjustment for concomitant variation.

It is usually pointless to measure the response (or to give the drug) in terms of a percentage or ratio relating the crude response

or dose to some such concomitant variable as body weight. Any influence such a factor may have is best measured subsequently by covariance analysis. Measurements of response based on such criteria as the mean percentage fall in blood sugar following the injection of insulin to rabbits are unnecessarily involved. A recalculation by Marks (unpublished) in which the results of a number of such tests were examined using only the final level of blood sugar at the second hour after injection gave rather more precise results than the original tests, if no correction were made in these latter for the correlation between initial blood sugar level and fall (series 1 and 2), and as precise a result in a test in which the correction was made (series 3), although no correction for initial blood sugar was made to the final blood sugar values in any comparison. The standard errors of the determination of potency were:

	Series 1	Series 2	Series 3
By percentage blood sugar reduction	0·049	0·075	0·030
By final blood sugar values	0·044	0·066	0·030

It would thus appear that much of the trouble to which the assayist goes in the determination of the potency of insulin by the rabbit method is unnecessary.

19.2. The choice between differences or final values

It seems to be almost universally assumed that, when both initial and final measurements of the responding tissue are possible in the same test object, it is preferable to employ the difference between them as the criterion of response rather than to use only the final values (on which one has to depend without choice in the many types of assay in which no initial measurements are possible). This preference must be based on the belief that more uniform results will be obtained by the elaboration of technique and that some degree of "animal variation" is eliminated by the procedure. Bound up with this assumption is the further one that if, as is very frequently the case, a correlation between the initial measurement and the subsequent change is demonstrable, it must be more accurate to use differences than to use final values, particularly if either an arbitrary correction or no correction for the initial levels is applied. In any assay employing differences between pairs of observations in which no correction is applied for the initial readings, an improvement in accuracy cannot be expected

compared with that attainable by the use of the final member of each pair alone, unless there is a sufficiently high correlation between members of pairs to compensate for the additional variance contributed by the initial measurement. The value which must be equalled or exceeded by the correlation coefficient concerned will depend on the variances of the two sets of measurements.

19.3. The regression equation

If the three variates, final level (Y), initial level (Z) and the dose, or any function of it, such as log dose (X), be so related that:

$$Y=a+b_{ZY.X}Z+b_{YX.Z}X$$

$b_{ZY.X}$ and $b_{YX.Z}$ are the partial regression coefficients, independent of X and Z respectively.

In tests with randomised animals there will be no correlation between X and Z, and the partial regression coefficients will be identical with the corresponding total regression coefficients. Thus, the same mean values for the bs will be found, if multiple regression is ignored. Then:

$$Y=a+b_{ZY}Z+b_{YX}X$$

$$Y-Z=a+(b_{ZY}-1)Z+b_{YX}X$$

Hence $(Y-Z)$ is linearly related to Z and X, and the slope b_{YX} is not affected by the change to $(Y-Z)$. Thus, in discussing the relative accuracy of assays based on Y or $(Y-Z)$, the effect on the slope of the regression line relating dose to response may be ignored. We are therefore mainly interested in the standard errors of $(Y-Z)$ and Y, which must be determined by trial. It will be preferable to choose the method of expressing the response which has the smaller standard error, since no correction, by covariance or otherwise, can be expected to reduce the variance of the other measure of response to a value below that found by applying the same type of correction to the measure exhibiting the smaller crude variance.

19.4. The characteristics of a satisfactory assay

Gaddum (*Biochem. J.*, **25**, 1,113, 1931) outlined some basic considerations which determined the choice of the best measure of response. They were:

i. A linear relationship over the widest possible range.

ii. A standard deviation which is independent of the response.

iii. A minimal value for the ratio $\frac{Ve}{b^2}$ (called λ^2 by Gaddum).

This ratio may not be minimal when *Ve* is minimal, but depends on the relationship between the two quantities, and a steep enough slope may compensate for the choice of a measure of response which does not have the smallest *Ve* of all available. These requirements have been enlarged upon by Bliss (cf. *Ind. Eng. & Chem.*, **13**, 84, 1941), who makes five points regarding a satisfactory assay:

i. Different samples of the same drug should show the same relative potencies in biological assay as under clinical test.

ii. On the co-ordinates used for biological assay the curve relating response to log dose should be a straight line and relatively steep when compared with the variation about the line. Either the curve should have been shown to have a constant, known slope by repeated tests over a considerable period of time or the slope should be determined as an integral part of each assay. Assumed relations between dose and effect are to be avoided.

iii. The potency of the unknown or sample should be determined by comparative test with a stable reference standard and expressed in units of this standard.

iv. The living material exposed to different doses of the standard and unknown must be as nearly equivalent as it can be made. Potential sources of variation, such as differences between individuals, litters, dates of treatment and sexes, should never coincide or be confounded with differences in treatment, but within these limitations the dosages and samples must be assigned at random. The analysis of variance or an equivalent technique should be used to segregate from the estimate of error the sources of variation that have been quarantined by the design of the assay. Variations in an initial measurement, such as the initial blood sugar in the rabbit insulin assay, or in a concomitant measurement, such as body weight, should be adjusted not by assumed relation, which is sometimes concealed in the definition of response, but rather from the internal evidence of self-contained experiments by covariance.

v. A determination of potency should always include an estimate of its error, computed as an integral part of the assay.

The first of Bliss's points applies, of course, to assays destined to determine the potency of therapeutic substances. It is not often possible, however, to be sure that this requirement is satisfied, as few drugs have been assayed in any real sense of the word, using human beings.

To the above should be added that the assay method should give reproducible results, which should be consistent with internal indications of precision, and that this precision should not alter violently from test to test or from time to time. If this occurs, the cause of the undesirable fluctuation must be sought in the procedure of the assay, when it may be found that apparently unimportant details, environmental or otherwise, may cause large changes in the mean square for error, in the slope of the dose-response line or in the effective doses. An extreme but most instructive example is the assay of sympathomimetic amines. Chance (*J. Pharmacol.*, **89**, 289, 1947) found that the toxicity of these substances to mice was influenced by the degree of hydration of the animal, the external temperature, the degree of confinement and the number of mice in a given space, as well as by such more usual factors as sex, body weight and strain differences. The influence of these various factors was such that more than a tenfold difference in median lethal dose could be produced for one compound by manipulation of the external environment alone.

The extent to which the design of tests may contribute to their precision has already been stressed. Up to five-fold or six-fold increases in precision have been reported as a result of segregating differences between individual animals in assays of coal-tar antipyretics with cats and parathyroid extract in dogs. Twofold to threefold increases in precision as the result of the elimination of differences between litters in the assay of vitamin D in rats were indicated by Bliss.

19.5. Transformations of the response

Transformations of the response may be employed to equalise variances, but are not always likely to be successful. Transforming the response to log response has been used successfully in assays involving direct enumeration, such as bacterial counts, as a means of obtaining a constant variance when the variance is a linear function of the square of the mean response in a group, but if such a transformation leads to a curved relationship between dose and response, nothing may have been gained. The analyst may choose to work over such a limited range of responses that curvature may be ignored, but this involves difficulties in planning the test and does not really solve the problem, as we then have a measure of low precision by which to assess the parallel nature of the lines relating

dose to response for the standard and the unknown, and a large error for the slope.

When an assay is well planned, so that the means of all responses to the standard and to the unknown are about equal and the same number of doses are given to all groups, quite large inequalities in variance at different dosage levels may be ignored with comparative safety, as there will be no bias in the assessment of potency. There may, however, be a bias in the estimation of mean squares associated with restrictions in design in the Latin square, and care must be exercised if this design is used. Again, when the variance changes with response, the regression line should in theory be fitted by the use of weights inversely proportional to the variance and proportional to the number of observations per group. This is again of little moment as long as the assay is so balanced that the mean of all responses to the standard is not much different from the mean of all responses to the unknown.

While every effort should be made to detect departures from linearity or from equality of variance, and to correct for them, it must be remembered that the numbers of observations made even in a series of tests may be insufficient to reveal significant departures, and that when really large numbers of observations have been made, it frequently turns out that both types of departure occur. Thus, in the absence of sufficient evidence to the contrary, we assume that a linear relationship holds and that variances can be pooled, but must bear in mind that it is rarely justifiable to attempt too much by way of statistical refinements when we are not even sure that our basic assumptions are sufficiently valid. It is much better to repeat an assay than to spend hours trying out various transformations or tests for the rejection of aberrant responses and so forth. These are better subjected to critical examination in the light of much evidence from repeated tests, when judgment can usually be passed without much doubt.

CHAPTER 20

THE RESPONSE LINEARLY RELATED TO THE DOSE

20.1. Introductory remarks

Recent work in which micro-organisms are grown in media containing all the substances necessary for growth except the substance being assayed has shown that the response, which may be a measurement of such factors as the acid production or turbidity of a culture, frequently bears a linear relationship to the dose, instead of to the logarithm of the dose. These *micro-biological assays* have been studied mathematically by Finney (*Quart. J. Pharm. Pharmacol.*, **18**, 77, 1945); Wood (*Analyst*, **71**, 1, 1946) and Wood and Finney (*Quart. J. Pharm. Pharmacol.*, **19**, 112, 1946).

Such an assay may be conducted along familiar lines, and, for instance, a four-point design as outlined in Chapter 10 may be used. If the ratio of the potency of the unknown to that of the standard is designated by R, then:

$$R=\frac{b_u}{b_s}$$

where the equations of the dose-response lines for the standard and unknown are:

$$Y=a_s+b_sX$$

and $$Y=a_u+b_uX$$

and b_u and b_s are respectively the slopes for the unknown and the standard preparations. In a four-point assay:

$$R=\frac{\bar{Y}_{u1}-\bar{Y}_{u2}}{\bar{Y}_{s1}-\bar{Y}_{s2}}$$

and $$VR=\frac{Ve}{d^2b_s^2}\left\{\frac{1}{n_{u1}}+\frac{1}{n_{u2}}+R^2\left(\frac{1}{n_{s1}}+\frac{1}{n_{s2}}\right)\right\}$$

where d=the dose interval for both substances and n is the number of observations on each dose level as indicated by the suffixes.

In assays with more than two doses of each substance the usual

tests of linearity may be applied, but there is, of course, no corresponding test for parallelism, as the slope of the regression line for the unknown will only equal that for the standard if R is unity.

20.2. Objections to the usual design

Unfortunately, the simple modification in the estimation of relative potency suggested above, although adequate for assays in which it is impossible to use the procedure we are about to discuss, is often insufficient to do justice to the possibilities of the test. For if it so happens that the linear dose-response relationship holds over a range of doses which extends as far as zero dose, the four-point or general $2\times n$-point design does not yield as much information as may be gained by distributing the same total number of observations in a different manner. Wood and Finney have shown, for example, that with a four-point assay, VR is a minimum when:

$$n_{u1}=n_{u2}=\frac{n_{s1}}{R}=\frac{n_{s2}}{R}=\frac{n}{2R+2}$$

where n=the total number of observations; whence:

$$VR_{\text{min.}}=\frac{(2R+2)^2}{nd^2}\cdot\frac{Ve}{b_s^2}$$

where R is not greater than unity.

Now, suppose that the n observations had been distributed between three dose-levels over as wide a range as possible within the limits of linearity of the dose-response curve, namely, zero dose, dose X of the standard and dose X' of the unknown, with n_o, n_s, and n_u, observations at each level respectively. Then, if $X=X'=2$ arbitrary units of each preparation:

$$VR=\frac{Ve}{4b_s^2}\left\{\frac{(1-R)^2}{n_o}+\frac{R^2}{n_s}+\frac{1}{n_u}\right\}$$

and VR is a minimum when:

$$\frac{n_o}{1-R}=\frac{n_s}{R}=n_u=\frac{n}{2}$$

whence:

$$VR_{\text{min.}}=\frac{1}{n}\cdot\frac{Ve}{b_s^2}$$

If we suppose that the doses in a four-point assay are as widely spaced as possible, d cannot exceed 2 (the two arbitrary units referred to above) and hence the variance of the three-point assay

is necessarily less than that of the four-point assay (and may be as little as one-quarter of it).

20.3. The "Common-Zero 5-point" design

In practice we would not prefer to use the three-point assay just described, because no test of linearity is available and no check on the similarity of the unknown to the standard is possible, except in so far as the unknown is seen to promote growth or whatever activity may be measured as a response. In a four-point assay in which the log dose is linearly related to response, parallel dose-response lines guarantee a specific degree of resemblance. Thus, a five-point assay is the least we should accept, as it provides a test of the linear response relationship on which the validity of the assay depends. In such an assay, VR depends on the distribution of observations over the different dose levels and on the particular levels chosen. A high, but not maximal level of precision in estimating R is combined with a high, but also not maximal precision in detecting departures from linearity if equal numbers of observations are made at all dosage levels, and if the doses of the standard and unknown are both 0, 1, 2 in arbitrary units (with the 0 common to both).

If, for example, 20 observations are to be made in total, four at each level, and the assay is to be considered satisfactory if R lies between 0·7 and 1·0 (the standard will always be administered in doses calculated to cover as nearly as possible the complete range of linear response), then VR, as estimated from the equation given below, will lie between:

$$\frac{0{\cdot}1.Ve}{b_s^2} \quad \text{and} \quad \frac{0{\cdot}0803.Ve}{b_s^2}$$

The corresponding VR for a distribution of five tubes per dosage level in a four-point assay with $d=2$, will lie between:

$$\frac{0{\cdot}2.Ve}{b_s^2} \quad \text{and} \quad \frac{0{\cdot}149.Ve}{b_s^2}$$

Hence R will always be less efficiently estimated if this design is employed, although not so efficiently estimated as if a three-point "common zero" assay were feasible.

The five-point assay with unequal distribution of observations among the five dosage levels can give a rather more precise estimate of R than if observations are equally distributed, but not only does it allow a less accurate test of departures from linearity, but also

the statistical estimation of R and VR becomes tedious, with relatively little gain in information. We shall therefore confine discussion to the balanced type of design. The general problem is discussed by Wood and Finney.

20.4. Estimating relative potency and its variance

The values assigned to b_s and b_u are not independent, since they share information from the common-zero group. The restraint put upon the dose-response lines, that they must intersect at $X=0$ (although not necessarily at $\bar{Y}_0$, the response at zero dose), is reflected in the equations determining them. The constants of the two regression lines—which must share a common value of a—are:

$$a=\bar{Y}_o-\frac{(L_s+L_u)}{7}$$

$$b_s=\frac{(\bar{Y}_{s2}-\bar{Y}_o)}{2}+\frac{(6L_u-L_s)}{70}$$

$$b_u=\frac{(\bar{Y}_{u2}-\bar{Y}_o)}{2}+\frac{(6L_s-L_u)}{70}$$

where $L_s=\bar{Y}_o+\bar{Y}_{s2}-2\bar{Y}_{s1}$ and $L_u=\bar{Y}_o+\bar{Y}_{u2}-2\bar{Y}_{u1}$.

Thus, unless $L_s=L_u=0$, a correction is introduced for the restraint. Note that the response to the middle dose of each preparation, $\bar{Y}_{s1}$ or $\bar{Y}_{u1}$ in this instance, does not normally influence the slope in a three-point line with balanced dosage groups, but that it does have some influence under present conditions.

The variance of R is given by the relationship:

$$VR=\frac{2Ve}{7nb_s^2}\cdot(8R^2-9R+8)$$

Ve, the error mean square, is calculated in the usual manner from the sum of squares *within* groups (including the "blank" at zero dose):

$$Ve=\frac{Sy_p^2}{n-5}$$

20.5. The test for linearity

The quantities L_s and L_u, defined above, measure the departures of the two dose-response lines from linearity. Their common variance, VL, is given by:

$$VL=\frac{30Ve}{n}$$

If either L_s or L_u significantly exceeds zero in a t-test, using $n-5$ degrees of freedom, the assay is not valid.

20.6. A practical example

Table 20.1 lists the protocols of an assay of the riboflavin content of a malt extract, using riboflavin as a standard. The assay is illustrated by Figure 20.1. Tubes containing a constant quantity

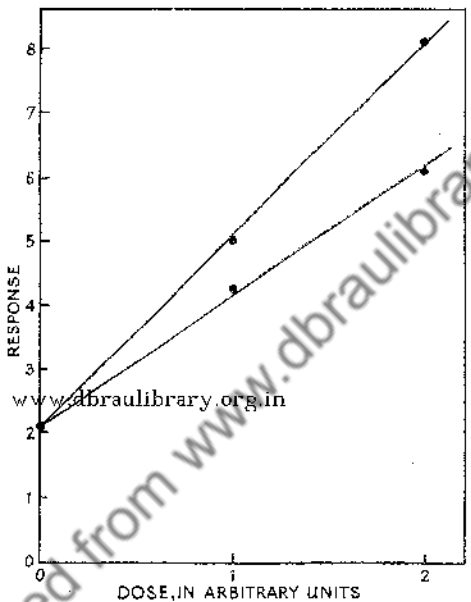


FIG. 20.1. The common zero five-point assay of Table 20.1.

of a basal medium inoculated with the micro-organism, *L. helveticus*, received in addition 0·0, 0·1 μg. (1 unit) and 0·2 μg. (2 units) of riboflavin and 0·025 gm. (1 unit) and 0·05 gm. (2 units) of malt (Wood, *Analyst*, **71**, 1, 1946). The response is measured by the millilitres of N/10 alkali required to neutralise the acid produced in each tube. Four tubes were used at each dosage level, 20 tubes in all. In this particular assay the basal medium itself contained a small quantity of riboflavin, so as to make the "zero" dose fall on the linear part of the dose-response curve, since the response at the extreme lower range was found not to be linearly related to the dose. There seems to be no valid objection to this procedure in the present type of test, although one can imagine circumstances in which it would be a hazardous practice.

TABLE 20.1

PROTOCOLS OF A TEST OF THE RIBOFLAVIN CONTENT OF MALT

(Adapted from Wood, *Analyst*, **71**, 1, 1946)

Dose (X)	ml. of N/10 acid produced in individual tubes (Y)				Means ($\bar{Y}_p$)
X_0 : Zero	1·90,	2·25,	2·00,	2·20	2·0875
X_{s1} : 1 unit riboflavin	4·85,	5·00,	5·25,	4·90	5·0000
X_{s2} : 2 units riboflavin	8·35,	8·20,	7·95,	7·80	8·0750
X_{u1} : 1 unit malt	4·00,	4·40,	4·50,	4·10	4·2500
X_{u2} : 2 units malt	6·05,	6·20,	6·10,	6·10	6·1125

$$L_s = 2{\cdot}0875 + 8{\cdot}0750 - 2 \times 5{\cdot}0000 = 0{\cdot}1625$$
$$L_u = 2{\cdot}0875 + 6{\cdot}1125 - 2 \times 4{\cdot}2500 = -0{\cdot}3000$$
$$a = 2{\cdot}0875 + \frac{0{\cdot}1375}{7} = 2{\cdot}1071$$
$$b_s = \frac{5{\cdot}9875}{2} - \frac{1{\cdot}9625}{70} = 2{\cdot}9657$$
$$b_u = \frac{4{\cdot}0250}{2} + \frac{1{\cdot}2750}{70} = 2{\cdot}0307$$
$$Sy_p^2 = 0{\cdot}46756$$

We determine the various quantities shown in Table 20.1 by the foregoing methods, whence:

$$R = \frac{b_u}{b_s} = \frac{2{\cdot}0307}{2{\cdot}9657} = 0{\cdot}6847$$

Hence the content of riboflavin in the malt is:

$$\frac{0{\cdot}2}{0{\cdot}05} \times 0{\cdot}6847 = 2{\cdot}74 \ \mu\text{g. per gm.}$$

The value of Ve is $\frac{0{\cdot}46756}{15} = 0{\cdot}03117$. Hence:

$$VR = \frac{2Ve(8 \times 0{\cdot}6847^2 - 9 \times 0{\cdot}6847 + 8)}{7 \times 20 \times 2{\cdot}9655^2}$$

and $s_R = 0{\cdot}01682$

Also, $VL = \frac{30Ve}{20} = 1{\cdot}5Ve$

$s_L = 0{\cdot}216$ (whence L_s and L_u do not significantly exceed zero).

The fiducial limits of R are given by:

$$R \pm ts_R$$

For $P=0{\cdot}05$, t with 15 degrees of freedom is 2·131; hence the fiducial limits of the result are:

$$(0{\cdot}6847 \pm 2{\cdot}131 \times 0{\cdot}0168)\ \frac{0{\cdot}2}{0{\cdot}05}\ \mu\text{g. per gm.}$$

$$= 2{\cdot}60 \text{ to } 2{\cdot}88\ \mu\text{g. per gm.}$$

In microbiological assays, there is usually very much less variation between replicates than in other types of assay, and the need for the calculation of exact fiducial limits of error should not usually arise, as the slopes are well determined. Fiducial limits for use in examples where this is not the case are discussed by Finney (reference above).

APPENDIX

NOTATION

X, the dose of a preparation given to an individual test object.

X_p, the dose given to a member of group p.

$\bar{X}_p$, the mean dose given to members of group p ($\bar{X}_p$ is normally equal to X_p).

$\bar{X}_s$, $\bar{X}_u$, the (weighted) mean of all doses of the standard and unknown respectively.

$\bar{X}$, the (weighted) mean of all doses of both standard and unknown.

x_p, the deviation of X_p from $\bar{X}_p$ (normally nil).

$\bar{x}_p$, the deviation of $\bar{X}_p$ from $\bar{X}$.

x, the deviation of X from $\bar{X}$.

Y, the response of a test object.

Y_p, the response of any member of group p, or in *quantal* assays the empirical probit for group p.

$\bar{Y}_p$, $\bar{Y}_s$, etc., are (weighted) means as above.

y_p, $\bar{y}_p$, etc., are deviations, as above.

T_p, the total of all responses in group p.

T_s, T_u, the total of all responses to the standard and unknown respectively.

T, the total of all responses to both standard and unknown.

W, a concomitant variable, such as the body weight of a test animal.

$\bar{W}_p$, $\bar{W}_s$, etc., are (weighted) means as above.

w_p, $\bar{w}_p$, etc., are deviations, as above. In *quantal* assays, w_p is the weight of the probit for group p. In Chapters 17 to 19, w is also used to denote any weight factor.

t_p, the total of all concomitant observations, W, in group p.

t_s, t_u, t, etc., the corresponding totals for W to T_s, T_u, T, etc.

n_p, the number of test objects in group p.

n_s, n_u, the total number of test objects receiving doses of the standard and unknown respectively.

n, the total number of test objects in an assay.

S denotes "the sum of . . .," thus:

SX = the sum of all values of X,

$Sn_p\bar{x}_p$ = the sum of all values of $(n_p \times \bar{x}_p)$.

SS denotes "the sum of the sums of . . .," thus:

SSX_p=the sum of the sums of all members of groups p, p', p'', etc.

SSx_py_p=the sum of the sums of all values of $(x_p \times y_p)$, for all groups, p, p', p'', etc.

V denotes a variance or mean square, thus:

Ve=the error mean square,

Vb=the variance of b.

s denotes a standard deviation or standard error, thus:

s alone=the square root of Ve.

s_b=the square root of Vb.

σ denotes a theoretical standard deviation or standard error, thus:

σ alone=the theoretical value (if any) of s,

σ_b=the theoretical value of s_b.

E, the value of Y as estimated from X.

R, the ratio of the potency of the unknown to that of the standard.

M, the logarithm of the ratio of the potency of the unknown to that of the standard.

b, the slope of a dose-response line, relating Y to X.

b_w, the slope of the line relating Y to W.

In factorial experiments:

I, the interval between log doses.

k, a polynomial or factorial coefficient.

B^2, the mean square associated with linear regression.

D^2, the mean square associated with differences between substances.

x, y, etc., are used as column headings in Tables describing factorial tests or tests using polynomial coefficients to indicate the type of sum or product being formed line by line in the Table.

While the above notation is standard throughout the present volume, it has been impossible to avoid the occasional use of the same symbol for different purposes. Thus, in Chapter 13.6, Y becomes the sum of a pair of observations and y their difference. Such departures are always described in the text and should lead to less confusion than the introduction of unfamiliar Greek or other symbols to denote functions of responses. The use of W and w to denote a concomitant variable and its deviations and of w as a weighting factor is not, perhaps, to be admired, but since these do not occur in the same context they have been allowed to stand.

PRINTED AND BOUND IN GREAT BRITAIN BY WILLIAM CLOWES AND SONS LTD, LONDON AND BECCLES