THE STATISTICS OF BIOASSAY

WITH SPECIAL REFERENCE TO THE VITAMINS

BY

C. I. BLISS

The Connecticut Agricultural Experiment Station and Yale University, New Haven, Connecticut

Reprinted, with additions, from
VITAMIN METHODS
Volume II

1952
ACADEMIC PRESS INC., PUBLISHERS
NEW YORK

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

In the last two decades, biological assay has engaged the attention of a growing number of statisticians and has been recognized increasingly by biologists. A general chapter on the statistical methods essential in vitamin research was included in the original plan for the collaborative volumes on *Vitamin Methods,* but its scope was soon limited by necessity to statistical methods in biological assay. Although it considered only techniques for vitamins, the resulting chapter became in effect a description of the common procedures in all areas of biological assay. Intended as a working manual for biologists, it gave the basic operating rules, even though some of them may be little more than useful approximations. In response to numerous requests, this chapter now appears independently.

We have used the opportunity of a separate printing to correct such misprints as have been discovered, to revise the original analysis of variance for slope-ratio assays, and to add an addendum of new material. The original page numbers have been retained through page 608, and this numbering is continued in the addendum. References to pages 41-275 refer to the article by Bliss and Gyorgy on "Animal Vitamin Assays" in Volume II of *Vitamin Methods,* from which one table and a figure have been reprinted in the addendum.

The original chapter would not have been possible without the generous aid of many friends. Thanks are due the following investigators for numerical data which appear in the examples: A. Black, P. R. Burkholder, E. W. Crampton, P. L. Harris, R. B. Hubbell, G. H. Kennedy, C. A. Morrell, B. L. Oser, L. I. Pugsley, L. Rosner, E. E. Snell, J. Waddell and C. L. Withner. The text has gained from the helpful criticisms of several colleagues, among them J. W. Tukey, W. G. Cochran, M. H. Quenouille, D. N. Nanda, Stuart Mudd, and the members of my class in biometry who checked the numerical examples. Finally, I am indebted to my secretary, Mrs. Anna Branchini, for her able assistance in the preparation of the original chapter, and to her successor, Theresa Santilli.

C. I. Bliss

June, 1952
New Haven, Conn.

CONTENTS

Statistical design and methods of analysis are essential in many phases of vitamin research. Some of the most important of these concern the sampling techniques for obtaining the original samples. The vitamin content of several vegetables, for example, has been shown to vary with the season, the variety, the region where grown, the type of

fertilizer, the method of storage, and the manner of cooking. Before characterizing a food with a single value, the experimenter should assure himself that variations from sources such as these are relatively unimportant. In other experiments the extent of vitamin deficiences may be measured in school children or an investigator may test the effect of adding known vitamins to the diet. Here again the success of the project depends upon a carefully controlled sampling plan or experimental design. Often the variation in vitamin content among subjects or samples is far greater than that in the method of measuring this content. Despite their importance, these variables are outside the scope of the present chapter.

The problem with which we are concerned here is the measurement of a given vitamin in a homogeneous sample of material. We will assume that in preparing it for testing, no appreciable amount of the vitamin is destroyed. We will also assume that interfering substances which would bias our measurements have been eliminated. Presumably the qualitative requirements in respect to basal media or diets have been solved satisfactorily. Our objective is to design the assay so that the variation in the living indicator will be minimized and so that we can assess the validity of the assay, obtain an estimate of the vitamin content, and measure the error of this estimate. The experimenter should realize, however, that other sources of error may be far greater than those inherent in the assay, large as the latter may be.

In the development of quantitative assays, the methods of analysis should be efficient in the sense that they use all of the relevant information in a given set of data. Such methods will be described in forms which facilitate their computation with a calculating machine. If a machine is available, even the most exact statistical techniques require only a small fraction of the time needed for the experimental phase of vitamin research. Where an assay technique is used routinely under stable conditions, valuable information is accumulated on certain of its essential characteristics. Short-cut or "inefficient" methods of estimation may then suffice and a few of these are considered. In a book on vitamin research, however, it seems preferable to give priority to the standard efficient techniques. A book of statistical tables is indispensable and by far the most useful of these is that by Fisher and Yates (1). Barlow's Tables (2) and a good table of logarithms are also invaluable.

Following a review of the principles which underlie biological determinations of potency, the present chapter will describe the statistical techniques of most value for biological assays of the vitamins, starting with the dosage-response curve and concluding with a section on multiple or repeated assays. It will not be possible to develop fully the biological

logic of each procedure and its mathematical basis not at all.* It is hoped, however, to indicate the purpose of each statistic and to provide working directions for its calculation. Wherever possible, each step is illustrated by a numerical example. In order to link the algebraic equations or formulas with their numerical solution, the examples are presented piecemeal. Each set of basic data is given a number and successive steps in its analysis a small letter, so that a subheading such as "Example 2c" introduces the third installment in the analysis of example 2. Similarly, alternate forms of an equation or formula are often given the original number followed by the letter a, b, etc. There is no standard nomenclature in statistics. The methods used here, however, are based primarily upon those described by that genius of modern statistics, R. A. Fisher, and much of the symbolism follows that in his standard text (3). A glossary of symbols is provided at the end of the chapter.

I. General Principles of Biological Assays

Biological assays have been considered at length in several articles and books (4–12) and their basic principles are now well established. As applied to the vitamins, the objective is to measure the biological potency of a preparation as it is reflected in some characteristic response. The responses which have been used are exemplified in detail in the chapter on animal assays and in the chapter of Volume I on the microbiological assays. A demonstration that a given preparation will cure a specific deficiency syndrome does not in itself constitute an assay. We need to determine how much of the biologically active material will produce what degree of cure. Among the quantitative tests of activity several types can be recognized, and a review of their characteristics and functions will introduce the detailed statistical procedures by which they are implemented.

1. Determinations of Activity

In designing a satisfactory assay, the first stage is to determine the relation between dosage and response with a single preparation. In each case the dose which reaches the site of action in the test organism is assumed to be proportional to that measured by the experimenter. As the dose of an active preparation is increased from that which gives no effect to one giving a maximum effect, the response increases (or decreases) continuously to form in many cases a sigmoid curve. Numerous examples of such dosage-response curves are given in the chapter on animal assays. No statistical techniques have been developed for handling a sigmoid curve which are as simple as those for a straight line.

* For these the reader is referred to the new comprehensive treatise on "Statistical Method in Biological Assay," by D. J. Finney (1952). (Note added in proof.)

In consequence, an initial objective is to select units of response and of dose which will lead to a straight line over a range that is wide enough for assay purposes. The units used in the calculation are sometimes referred to as "metameters" (13), and the problem arises of selecting both a response metameter and a dose metameter, usually on empirical grounds.

The response metameter varies considerably from one assay to another, but only two dose metameters lead to satisfactory assays. The more important of these is the logarithm of the dose, and most of this chapter concerns responses which, within limits, are substantially linear when plotted against the log dose. These include most of the animal assays and some microbiological assays. The other dose metameter is the arithmetic dose (x) raised to the i^{th} power or x^i. If $i = 1$, then the response can be plotted as a straight line against the original units of dose over an adequate range. Some curves which are otherwise curvilinear, however, can be straightened by plotting the response metameter against x raised to a power either smaller or larger than one. Many microbiological assays fall into this general category as well as several of the animal assays. This type of dosage-response curve leads to the so-called slope-ratio assays, which are considered later in the present chapter. In either case a single dosage-response curve is computed with the same equations, whether the dose is expressed in terms of logarithms or in terms of x^i.

Dosage-response curves serve several functions. They indicate the dosage range which is useful for biological assays. The quantitative importance of suspected sources of variation can be examined in the same experiments and procedures developed for minimizing their interference in later determinations of potency. The computed results can also be used for estimating the inherent precision of a given assay technique. For these and other reasons a quantitative knowledge of the dosage-response relation is an essential prerequisite to the design of an effective assay. These initial problems are considered in sections II and III of the present chapter.

2. Comparative Biological Assays

Comparative biological assays require two or more preparations or compounds, one of which may be designated as the "standard" (S) and the other as the "unknown" (U). The potency of the unknown relative to the standard is determined in respect to some specific activity, whether or not it is known to differ chemically from the standard. The purpose of the assay is to determine its relative activity under a given set of con-

ditions. As additional requirements to those for the dosage-response curve, the response to both the standard and the unknown must be measurable in the same units and the two preparations must be compared within tests.

At the start of the assay the unknown is assigned an assumed potency in units of the standard, and both preparations are given at the same dosage levels. It is preferable that the mean response to these comparable doses of the standard and of the unknown should not differ significantly. The underlying relation between dosage and response may differ quantitatively from one laboratory to another and from one experiment to another in the same laboratory, so that it should be redetermined as an integral part of each assay. This requires a minimum of two dosage levels of either the standard or the unknown.

For an adequate assay there should be at least two dosage levels of both preparations. It is then possible to determine whether the relative potency is independent of the level of response. When the dose metameter is the log dose, this is equivalent to determining whether the curves computed separately for the standard and for the unknown have the same slope within the sampling error. When the response plots linearly against arithmetic dosage units or against x^4, we have a slope-ratio assay and test instead whether the two dosage-response lines intersect at zero dose. Assays of the first type with a graded response are considered in section IV of the present chapter, covariance methods for improving their precision in section V, and those with an all-or-none response in section VI. Slope-ratio assays are described in section VII.

In each case an estimate of the inherent precision of the assay is considered as important as that of the observed potency. Precision is expressed in terms of the standard error or of the confidence or fiducial limits as computed from the variation in response within a single assay. The assayed potency in repeated tests of a single unknown will rarely show significantly better agreement than would be predicted from the internal evidence of each assay and may not agree as well. This depends upon the relative importance of sources of variation within and between assays and can be determined only by experimental trial. Methods for separating these two components from the data of repeated assays of the same unknown are discussed in section VIII of this chapter.

3. Analytical Biological Assays

When the potency of a vitamin preparation is standardized biologically and in certain research problems, analytical biological assays are indicated. In addition to the requirements of the two preceding cate-

gories, the unknown is then assumed to differ from the standard only in its concentration in a completely inert diluent. Unlike a comparative biological assay both the test and the standard preparations in an analytical biological assay are assumed to be so nearly identical chemically that no matter how they are administered nor what the organism, their relative potency will be identical within the experimental error. The objective of the experiment is to determine how many units of the standard vitamin are present in a given quantity of the unknown.

The design of an analytical assay is essentially the same as that of a comparative biological assay but in view of the assumption of qualitative equivalence it should be unnecessary to specify in exact detail just how the assay is conducted. Yet most official assays require only one technique which is described in meticulous detail. This suggests that in practice acceptable test preparations may differ qualitatively from the reference standard or be dissolved in a carrier which is not wholly inert, so that quantitative comparisons are possible only when the method of assay is completely specified.

Even though it is required theoretically that the relative potency must be identical with all possible techniques, most analytical assays are based upon a single procedure which is repeated routinely in a given laboratory. Under stable test conditions, valuable information is accumulated on certain essential assay characteristics, such as the slope of the dosage-response curve and the standard deviation about this curve. Methods for using this information to improve the precision of a current assay are discussed in the last section of the present chapter.

4. Inspection Assays

Many biological assays, most notably those for vitamins A and D, have the sole purpose of determining whether the test preparation contains the claimed amount of the vitamin. These pass-or-fail tests may be carried out in the producer's laboratory or by a regulatory agency. It is not necessary in such cases to find out how much of the vitamin is present, as in the preceding two sections, but only whether the amount claimed is there.

In general, these official assays leave much to be desired. In addition to the assumptions in the preceding categories, adequate inspection assays would require agreement as to consumers' and producers' risks and adoption of a fixed method of inspection. Before pass-fail criteria can be established, the assay procedure must be standardized in each laboratory, so that the process is "in control," as the term is used by the statisticians in quality control. One of these techniques is the quality

control chart described in section VIII of the present chapter. Instead of considering each assay individually and only upon its own merits, emphasis is shifted to the inspection process, which must meet certain standards over a continuous series of assays. Then, taking into account the risk to the consumer of his purchasing substandard preparations and the risk to the producer of having good lots rejected unfairly, consumers' and producers' risks need to be balanced by mutual agreement. At the present time no vitamin biological assays have been put upon this basis, nor indeed have the biological assays of other biologically active preparations. This is a development which still lies in the future and cannot be considered further in the present chapter.

II. The Dosage-Response Curve and Its Error

The relation between dosage and response is typically a sigmoid curve. If the dose of vitamin can be increased sufficiently, a stage is reached beyond which further increases in the dose do not increase the response. As the dose is decreased, the response reaches a lower limit, not abruptly but usually through some transitional phase. This lower limit is measured by the negative controls. In experiments with a graded response only the central portion of the curve is used, within which the relation between dosage and response is essentially linear. In most animal vitamin assays the central portion covers the widest range when the dose is expressed in logarithms, and this is the relation which will be considered in the present section. The same statistical methods are used in computing the curve when the dose is expressed in arithmetic units and both types are illustrated in the chapter on animal vitamin assays.

The initial step is to select a response metameter which will plot as a straight line against the log dose of vitamin. The computing labor is greatly reduced if at the same time the individual observations at each dosage level are equally variable. The variability in the response metameter is assumed further to be distributed in reasonable agreement with the normal or Gaussian curve, a basic random distribution of first importance in statistical theory. In contrast, each recorded dose is presumed to be free of variation and without sampling error. Occasionally two of these requirements are mutually exclusive in that the response metameter which plots linearly against the log dose has a variability which is related to the response, as in an all-or-none reaction. These special cases will be considered in a later section. Here we are concerned with the more usual situation where each individual contributes equally to the determination of the curve.

In examining the relation to dose, a single preparation of vitamin is administered at several dosage levels, preferably in a logarithmic series.

The first stage in the analysis is graphic with cross-section paper. The response is plotted on the ordinate against the log dose on the abscissa and the plotted points are examined for their agreement with a straight line and equal scatter about this line. If they describe a smooth curve, another function of the response is tried. If they seem satisfactorily linear a line is fitted by inspection with the edge of a transparent rule or triangle, excluding any doses at the end of the curve where the response seems clearly to be restricted by an upper or a lower limit. The next stage is to compute the straight line which best fits the responses in the linear zone of log doses. The calculation not only adjusts objectively for the variability of individual points which visually may be given too much or too little weight, but it also enables the experimenter to measure quantitatively the variation about the line and to estimate the precision of a given assay technique.

1. The Calculation of the Line

The straight line relating the response metameter to the log dose is computed by least squares, so that the line which relates y to the log dose x is that which minimizes the sums of the squares of the deviations in y. If so wide a range of doses has been used that the response at the largest dose falls near the maximum which is possible or the "ceiling" or that at the lowest dose near the minimum or the "floor," it may be necessary to omit the data for one or more doses at one or both ends of the series, as is illustrated in Figs. 17, 18A, B, 19A, and 23A, B of the chapter on animal assay. This elimination requires considerable judgment and can be done most satisfactorily by inspection. The line is fitted only to the points in the intermediate linear zone.

The calculation of a dosage-response curve by least squares consists of computing the statistics a and b in the equation of a straight line

$$Y = a + b(X - \bar{x}) \tag{1}$$

where a is equal numerically to the mean response, $\bar{y}$, over all dosage levels included in the calculation, $\bar{x}$ is the mean log dose, and b is the slope of the line. The small letters x and y refer respectively to the observed log dose and the observed response metameter, referred to generically as "variates," and the capital letter Y to the response predicted by the equation for any contemplated log dose X. The entire equation is commonly known as the regression equation and its slope b as the regression coefficient.

The statistics a and b are computed from a given set of observations and in repeated assays their fluctuations are independent of each other. These observations form only one sample from a potentially unlimited

control chart described in section VIII of the present chapter. Instead of considering each assay individually and only upon its own merits, emphasis is shifted to the inspection process, which must meet certain standards over a continuous series of assays. Then, taking into account the risk to the consumer of his purchasing substandard preparations and the risk to the producer of having good lots rejected unfairly, consumers' and producers' risks need to be balanced by mutual agreement. At the present time no vitamin biological assays have been put upon this basis, nor indeed have the biological assays of other biologically active preparations. This is a development which still lies in the future and cannot be considered further in the present chapter.

II. The Dosage-Response Curve and Its Error

The relation between dosage and response is typically a sigmoid curve. If the dose of vitamin can be increased sufficiently, a stage is reached beyond which further increases in the dose do not increase the response. As the dose is decreased, the response reaches a lower limit, not abruptly but usually through some transitional phase. This lower limit is measured by the negative controls. In experiments with a graded response only the central portion of the curve is used, within which the relation between dosage and response is essentially linear. In most animal vitamin assays the central portion covers the widest range when the dose is expressed in logarithms, and this is the relation which will be considered in the present section. The same statistical methods are used in computing the curve when the dose is expressed in arithmetic units and both types are illustrated in the chapter on animal vitamin assays.

The initial step is to select a response metameter which will plot as a straight line against the log dose of vitamin. The computing labor is greatly reduced if at the same time the individual observations at each dosage level are equally variable. The variability in the response metameter is assumed further to be distributed in reasonable agreement with the normal or Gaussian curve, a basic random distribution of first importance in statistical theory. In contrast, each recorded dose is presumed to be free of variation and without sampling error. Occasionally two of these requirements are mutually exclusive in that the response metameter which plots linearly against the log dose has a variability which is related to the response, as in an all-or-none reaction. These special cases will be considered in a later section. Here we are concerned with the more usual situation where each individual contributes equally to the determination of the curve.

In examining the relation to dose, a single preparation of vitamin is administered at several dosage levels, preferably in a logarithmic series.

The first stage in the analysis is graphic with cross-section paper. The response is plotted on the ordinate against the log dose on the abscissa and the plotted points are examined for their agreement with a straight line and equal scatter about this line. If they describe a smooth curve, another function of the response is tried. If they seem satisfactorily linear a line is fitted by inspection with the edge of a transparent rule or triangle, excluding any doses at the end of the curve where the response seems clearly to be restricted by an upper or a lower limit. The next stage is to compute the straight line which best fits the responses in the linear zone of log doses. The calculation not only adjusts objectively for the variability of individual points which visually may be given too much or too little weight, but it also enables the experimenter to measure quantitatively the variation about the line and to estimate the precision of a given assay technique.

1. The Calculation of the Line

The straight line relating the response metameter to the log dose is computed by least squares, so that the line which relates y to the log dose x is that which minimizes the sums of the squares of the deviations in y. If so wide a range of doses has been used that the response at the largest dose falls near the maximum which is possible or the "ceiling" or that at the lowest dose near the minimum or the "floor," it may be necessary to omit the data for one or more doses at one or both ends of the series, as is illustrated in Figs. 17, 18A, B, 19A, and 23A, B of the chapter on animal assay. This elimination requires considerable judgment and can be done most satisfactorily by inspection. The line is fitted only to the points in the intermediate linear zone.

The calculation of a dosage-response curve by least squares consists of computing the statistics a and b in the equation of a straight line

$$Y = a + b(X - \bar{x}) \tag{1}$$

where a is equal numerically to the mean response, $\bar{y}$, over all dosage levels included in the calculation, $\bar{x}$ is the mean log dose, and b is the slope of the line. The small letters x and y refer respectively to the observed log dose and the observed response metameter, referred to generically as "variates," and the capital letter Y to the response predicted by the equation for any contemplated log dose X. The entire equation is commonly known as the regression equation and its slope b as the regression coefficient.

The statistics a and b are computed from a given set of observations and in repeated assays their fluctuations are independent of each other. These observations form only one sample from a potentially unlimited

number of additional sets of similar observations which might be made under exactly the same conditions. Each additional set could be characterized by its own a and b, none duplicating any other. In the aggregate we might accumulate theoretically a population of such a's and b's. The central values about which the individual estimates would cluster if we continued the sampling indefinitely are known as the *parameters* of the population and are designated by the Greek letters α and β respectively. The statistics a and b when calculated by least squares are unbiased estimates of the parameters α and β. The form of equation which is most convenient for computing the a and b of a particular dosage-response curve will vary with the design of the experiment as described in the following three sections.

A. COMPUTATION FROM INDIVIDUAL OBSERVATIONS

The most general case is that in which the dosage level for each observation is determined separately, so that there are relatively few dosage levels with two or more responses and no regular sequence between successive doses. Each known log dose x is paired with an observed response y and from the values a line is computed. The position of this line is such that at $\bar{x}$, the mean of the x's, it will pass through $\bar{y}$, the mean of the y's. The equations for the means are

$$\bar{x} = \frac{S(x)}{N} \tag{2}$$

and

$$\bar{y} = \frac{S(y)}{N} \tag{3}$$

where $S(\ )$ is the sum of all values of the variate included in the parentheses and N is the total number of observations.

The slope or tangent of the line depends upon two terms. We first compute the sum of the squares of the deviations from the mean log dose as

$$[x^2] = S(x - \bar{x})^2 = S(x^2) - \frac{S^2(x)}{N} \tag{4}$$

In this and all other equations square brackets [ ] enclose the sum of deviations from the mean of each variate. The first form of the above equation is the easier to understand, but the second is much easier to compute. Algebraically they are identical. The sum of the products of x and y, as measured from their respective means, is computed next as

$$[xy] = S\left\{(x - \bar{x})(y - \bar{y})\right\} = S(xy) - \frac{S(x)S(y)}{N} \tag{5}$$

The sum of squares $[x^2]$ is invariably a positive number, but the sum of products $[xy]$ may be either positive or negative, depending upon whether the response is related directly or inversely to the log dose. The ratio of these two terms is the slope of the dosage-response curve or

$$b = \frac{[xy]}{[x^2]} \tag{6}$$

The slope has the same sign, of course, as $[xy]$.

It is now possible to plot the straight line of Eq. 1. For this purpose it is convenient to condense the equation to the form

$$Y = a' + bX \tag{1a}$$

where $a' = \bar{y} - b\bar{x}$. The expected response Y is computed for two contemplated values of X, one at the lower end and the other at the upper end of the range of doses. These are plotted on the diagram and connected by a straight line.

Example 1a. The calculation may be illustrated by the dosage-response curve for a natural source of vitamin K (dried pig's liver) as measured in chicks by Schønheyder (14). The greater the dose of vitamin K fed for three days to a depleted chick, the smaller is the concentration of clotting agent required to clot its plasma in three minutes. The response metameter y is the log concentration of clotting agent and the dose metameter x the log (mg./g. chick/day). When y is plotted against x the data can be fitted by a straight line (Fig. 14,A, chapter on animal assay). The results for 15 chicks and the calculation of the dosage-response curve from these values are given in Table I. For plotting, $X = 0.2$ and 1.2 may be substituted in Eq. 1a to obtain $Y = 2.63$ and 0.74 respectively.

TABLE I. Dosage-Response Curve relating the Log concentration of Clotting agent y to the log dose x of natural Vitamin K in Individual Chicks

Data of F. Schønheyder (14)

x	y	x	y	x	y
.20	2.70	.64	1.67	.91	1.28
.34	2.21	.68	1.79	1.00	1.08
.45	2.25	.78	1.59	1.01	1.00
.48	2.13	.83	1.60	1.05	.95
.57	1.89	.86	1.32	1.17	.90

$S(x) = 10.97, \quad S(y) = 24.36, \quad N = 15$

$\bar{x} = 10.97/15 = .7313$ (Eq. 2); $\bar{y} = 24.36/15 = 1.6240$ (Eq. 3)

$[x^2] = 9.1419 - 10.97^2/15 = 1.1192$ (Eq. 4)

$[xy] = 15.6983 - 10.97 \times 24.36/15 = -2.1170$ (Eq. 5)

$b = -2.1170/1.1192 = -1.8915$ (Eq. 6)

$Y = 1.6240 - 1.8915(X - .7313)$ (Eq. 1)

or $a' = 1.6240 + 1.8915 \times .7313 = 3.0073$

giving $Y = 3.0073 - 1.8915X$ (Eq. 1a).

B. COMPUTATION WITH SEVERAL RESPONSES AT EACH DOSE

In measuring the relation between the dosage of vitamin and the response, several observations are usually made at each dosage level. The calculation can be shortened by totaling the y's at each dose to obtain the "dose total" T_d and computing

$$\bar{x} = \frac{S(fx)}{N} \tag{2a}$$

and

$$\bar{y} = \frac{S(T_d)}{N} \tag{3a}$$

where f is the frequency or the number of individual observations at a given dose x and $N = S(f)$. The denominator of the slope is determined as

$$[x^2] = S(fx^2) - \frac{S^2(fx)}{N} \tag{4a}$$

and its numerator or the sum of products as

$$[xy] = S(xT_d) - \frac{S(fx)S(T_d)}{N} \tag{5a}$$

Alternatively, if the number of observations (f) is constant at each of the k dosage levels, $fS(x) = S(fx)$ and

$$\bar{x} = \frac{fS(x)}{N} = \frac{S(x)}{k} \tag{2b}$$

$$[x^2] = f\left\{S(x^2) - \frac{S^2(x)}{k}\right\} \tag{4b}$$

and

$$[xy] = S(xT_d) - \frac{fS(x)S(T_d)}{N} \tag{5b}$$

Example 2a. The above procedures may be illustrated by the bradycardia technique for the assay of vitamin B_1 as reported by Leong (15). The duration of cure of bradycardia in depleted rats following a single dose of vitamin B_1 was measured to the nearest half day as shown in the left side of Table II. It is apparent from Fig. 1 that in terms of log days as given in the right side of the table the response varied uniformly at all dosage levels. To shorten the arithmetic the characteristic 1 has been subtracted from each log dose, but if desired, the equation of the computed line (Eq. 1) can be decoded by adding 1 to $\bar{x}$. In this range of doses there was no evidence of departure from a linear trend. The dosage-response curve was calculated from the logarithms in the right

TABLE II. Dosage-Response Curve for Vitamin B_1 (International Standard) by Length of Cure of Bradycardia in Rats

Data of P. C. Leong (15)

Dose mg.	Length of cure in days in individual rats					Log dose (−1), x	Response in log days for each rat, y					T_d
10	2	2	2.5	3	3	.000	.30	.30	.40	.48	.48	1.96
20	3.5	4.5	4.5	5	5.5	.301	.54	.65	.65	.70	.74	3.28
30	4.5	5.5	6	6	8	.477	.65	.74	.78	.78	.90	3.85
40	6	7	7.5	9	10	.602	.78	.85	.88	.95	1.00	4.46
Totals						$S(x) = 1.380$						$S(T_d) = 13.55$

$f = 5$ at all doses, $k = 4$, and $N = 20$

$\bar{x} = 1.380/4 = .3450$ (Eq. 2b); $\bar{y} = 13.55/20 = .6775$ (Eq. 3a)

$[x^2] = 5\{.680534 - 1.380^2/4\} = 1.0222$ (Eq. 4b)

$[xy] = 5.50865 - 5 \times 1.380 \times 13.55/20 = .8339$ (Eq. 5b)

$b = .8339/1.0222 = .8158$ (Eq. 6)

$Y = .6775 + .8158(X - .345)$ (Eq. 1)

or $Y = .3960 + .8158X$ (Eq. 1a), since $a' = .6775 - .8158 \times .345 = .3960$

of Table II as shown below it in detail. The computed line has been plotted in Fig. 1.

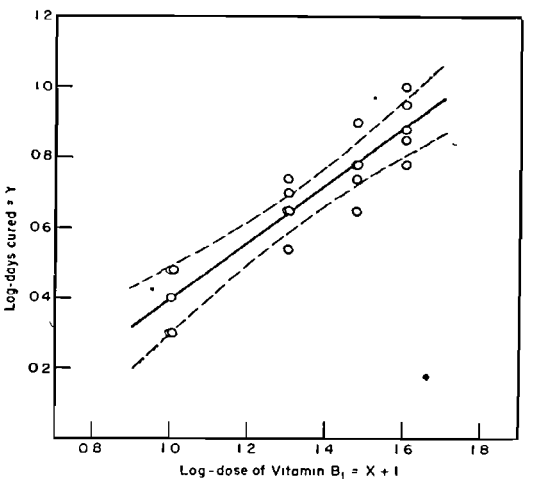

FIG. 1. Dosage-response curve relating length of cure from bradycardia in rats to log dose of vitamin B_1, from data in Table II.

C. COMPUTATION WITH CODED DOSES

The above equations do not depend upon the spacing of successive doses. However, if the experimenter spaces his doses equally on a logarithmic scale and assigns the same number of individuals to each

dosage level, he can save himself a great deal of time in the computation. Moreover, as will be shown later, it is then much easier to test the linearity of the curve. Perhaps the most frequent scale of doses is that for which the log interval $i = .3010$, which is obtained by multiplying (or dividing) successive doses by 2, i.e. 1, 2, 4, 8 · · · . If this gives too coarse a scale, $\sqrt{2}$ may be substituted, for which $i = .1505$. A convenient series with $i = 0.1505$ is the following: 4.25, 6, 8.5, 12, 17, 24, 34, 48, 68. A more open scale with $i = .1761$ is given by the series: 3.6, 5.3, 8, 12, 18, 27, 40.5, 60.8.

With a balanced design the observed log doses can be replaced provisionally by systems of small whole numbers which have the property of totaling 0. Any doses at the ends of the series, beyond the range that can be fitted by a straight line, are first eliminated. Each dose is then coded by a system which depends upon whether the number of doses is odd or even. If odd, the middle dose is assigned the value of 0 and the doses are numbered consecutively above and below 0 as $x_1 = 1, 2, 3 \cdot \cdot \cdot$, and $-1, -2, -3 \cdot \cdot \cdot$. Given an even number of doses those in the upper half are numbered $x_1 = 1, 3, 5 \cdot \cdot \cdot$, and in the lower half $x_1 = -1, -3, -5 \cdot \cdot \cdot$, so that the interval between successive coefficients is 2. The sums of squares and products are computed with the x_1's and converted to units of log dose by a factor in the equation for the slope.

The slope measuring the change in y for each unit increase in the log dose is computed by the equation

$$b = \frac{S(x_1 T_d)}{I' f S(x_1^2)} \tag{7}$$

where f is the number of observations at each x_1 and I' is the ratio of the interval i (in logarithms or other units) between successive doses of vitamin to the interval between the coefficients x_1. If the number of doses is odd, $I' = i$; if even, $I' = i/2$. Equation 7 with $i = 1$ is also used in computing the growth response of a rat or chick from weekly weighing as in the response to vitamin A in example 4a. For plotting in terms of the coded x_1 the equation for b simplifies to

$$b_1 = \frac{S(x_1 T_d)}{f S(x_1^2)} \tag{7a}$$

The slope may also be computed by Eq. 6, in which case

$$[x^2] = I'^2 f S(x_1^2) \tag{4c}$$

and

$$[xy] = I' S(x_1 T_d) \tag{5c}$$

The mean log dose $\bar{x}$ is the log dose corresponding to $x_1 = 0$, if the number of doses is odd, or the average of the log doses for $x_1 = -1$ and $x_1 = 1$ if the number is even. The mean response $\bar{y}$ is computed by Eq. 3a as before.

Example 3a. The calculation of a balanced experiment may be illustrated by the dosage-response curve for niacin from the growth of *Lactobacillus arabinosus*. The doses represent a logarithmically spaced series in which each dose was multiplied by $\sqrt{2}$ to obtain the next higher dose. The doses are shown in the left column of Table III, each dose being replicated in three tubes. Following an incubation period the

TABLE III. Dosage-Response Curve for Niacin in Terms of the Titer Measuring the Growth of *Lactobacillus arabinosus* in 72 Hours

Data of P. R. Burkholder

Dose μg./ tube	Titer in three replicates			Coded dose x_1	Log(titer — 1.10) (+ 1) y			Total T_d	Mean $\bar{y}_d$	Quadratic coefficients x_2
0	1.10	1.08	1.13							
.0177	1.66	1.72	1.53		.75	.79	.63	2.17	.72	
.0250	1.71	1.78	1.70	— 4	.79	.83	.78	2.40	.80	28
.0354	1.98	1.95	1.98	— 3	.94	.93	.94	2.81	.94	7
.0500	2.23	2.39	2.31	— 2	1.05	1.11	1.08	3.24	1.08	— 8
.0707	2.68	2.68	2.60	— 1	1.20	1.20	1.18	3.58	1.19	— 17
.1000	3.02	3.23	3.23	0	1.28	1.33	1.33	3.94	1.31	— 20
.1414	4.07	4.03	4.00	1	1.47	1.47	1.46	4.40	1.47	— 17
.2000	5.27	5.15	5.10	2	1.62	1.61	1.60	4.83	1.61	— 8
.2828	6.74	6.73	6.74	3	1.75	1.75	1.75	5.25	1.75	7
.4000	8.45	8.35	8.38	4	1.87	1.86	1.86	5.59	1.86	28
.5657	9.78	9.86	9.73		1.94	1.94	1.94	5.82	1.94	
.8000	10.70	10.84	10.76		1.98	1.99	1.98	5.95	1.98	
1.1314	10.92	10.89	10.56		1.99	1.99	1.98	5.96	1.99	
1.6000	10.94	11.13	11.17		1.99	2.00	2.00	5.99	2.00	
				$S(x_1) = 0$				$S(T_d) = 36.04$		$S(x_2) = 0$

$f = 3$, $k = 9$, $N = 3 \times 9 = 27$, $I' = i = .1505$

$I'fS(x_1^2) = .1505 \times 3 \times 60 = 27.09$, $S(x_1T_d) = 24.08$, $b = 24.08/27.09 = .8889$ (Eq. 7)

$\bar{x} =$ log dose at $x_1 = 0$ or $\bar{x} =$ log (0.1000) $= -1.000$

$\bar{y} = 36.04/27 = 1.3348$ (Eq. 3a); $Y = 1.3348 + .8889(X + 1.000)$ (Eq. 1)

For plotting in Fig. 2, $b_1 = 24.08/180 = .1338$ (Eq. 7a), and $Y = 1.3348 + .1338x_1$.

organisms were killed by heat and each tube titrated with 0.1 N NaOH to obtain the titers shown at the left of the table. These were changed to

response metameters by first subtracting the mean titer for the three negative controls from the titer for each other tube and converting the differences to logarithms. The means of the three replicates $\bar{y}_d$ were then plotted against equal dosage intervals along the abscissa in Fig. 2. Des-

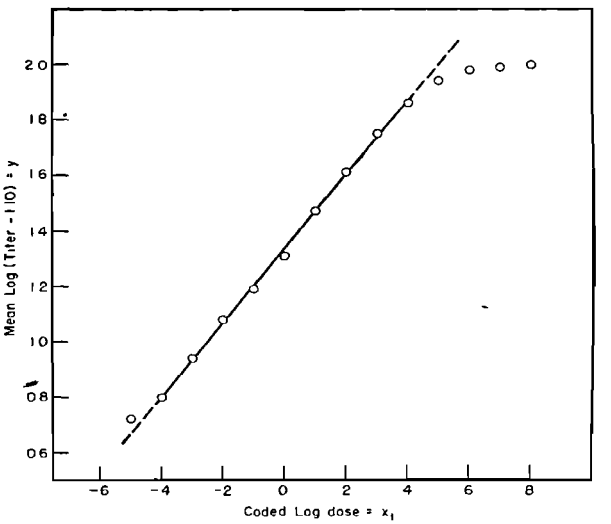

FIG. 2. Dosage-response curve for microbiological assay of niacin, showing selection of approximately linear zone of response. Data from Table III.

pite the sigmoid character of the curve as a whole, the response to the nine doses from 0.0250 μg.[1] to 0.4000 μg. was substantially linear and has been fitted with a straight line. The steps in the calculation are shown in full below Table III and the resulting straight line in terms of x_1 has been plotted in Fig. 2. Within this dosage range assays could be based on the assumption of a linear relation between log dose and the response metameter.

2. Analysis of the Variation about the Line

Given the equation for the dosage-response curve in an experiment, the response Y can be computed for any contemplated log dose X in the range covered by the observations. The predictive value of the curve,

[1] μg. = microgram.

however, depends upon how closely the observed responses agree with those expected from the equation of the line. For judging the potential value of a method as a biological assay, some measure of the variability about the line is as essential a characteristic as the equation itself. This variation may be due to many causes, some of which can be identified in a well-planned experiment. A method is needed for separating the latter components from the background of random variation due to unidentified causes. At the same time the calculations must lead to an estimate of the variability about the curve, principally in terms of the random components.

The major statistical tool for this and many other purposes is the *analysis of variance,* so-called because the unit of calculation is the variance or the square of the deviation of a single observation from a mean. Variances have the advantage of being additive, so that they lead to an unusually simple mode of calculation. Moreover, the variation accounted for by terms such as the slope of the line or by differences in statistics, such as between means, can be expressed in the same basic units. The analysis of variance has been described in many statistical texts and recently in three papers which consider the assumptions underlying the analysis of variance (16), pitfalls in its use (17) and transformations for extending its applicability (18), to which the reader is referred. In the present chapter it will be used initially for analyzing and estimating the variation about the dosage-response line and in later sections for solving a wide variety of assay problems.

A. THE STANDARD DEVIATION ABOUT THE COMPUTED LINE

The simplest dosage-response curve is one in which there is a single response at each of the several dosage levels, numbering N in all. In fitting a curve two statistics are determined from the y's, the mean $\bar{y}$ from the responses at all dosage levels and a slope b from the independent paired values of y and x. Let us first consider the variation about $\bar{y}$. If we square and sum all the y's, regardless of x, we obtain a "sum of squares" $S(y^2)$ as measured from zero, which may be of arbitrary origin if the responses have been coded. No restriction has been imposed on the freedom of the y's to vary and the sum has N degrees of freedom or as many as there are values of y. However, if we subtract $\bar{y}$ from each y, the N differences have $N - 1$ degrees of freedom, and $\bar{y}$ accounts for the one that is lost. When $N - 1$ such differences are added in any order the last or N^{th} difference is the value which makes all N differences total zero; it cannot be anything else and hence has no degree of freedom. If the N differences are squared before they are added, the process of

squaring does not restore the lost degree of freedom, and in consequence the sum of squares of the differences has $N - 1$ degrees of freedom. Corresponding to this loss, the sum of squares of the differences is less than the sum of squares of the original y's.

In this way a sum of squares of deviations from the mean with $N - 1$ degrees of freedom can be isolated from the sum of squares of the original y's. The difference between the two sums is a sum of squares with one degree of freedom for the difference between the mean and zero. Given any two of these values, the third can be obtained by addition or subtraction. In practice it is easiest to compute the sum of squares for the difference between the mean and zero as $S^2(y)/N$, which we will usually call the "correction for the mean," and next in order $S(y^2)$, the crude sum of squares of the original values. For this reason $[y^2]$, the sum of squares of the deviations from the mean $S(y - \bar{y})^2$, is computed indirectly as the difference between the other two terms or

$$[y^2] = S(y^2) - \frac{S^2(y)}{N} \tag{8}$$

which is analogous to $[x^2]$ as computed by Eq. 4. In the analysis of variance, the full term "sum of squared deviations" is abbreviated to "sum of squares" and $[y^2]$ is termed the "total sum of squares," since there we are concerned with deviations from statistics based upon the original values.

Part of the sum of squares about $\bar{y}$, however, is due to a variation in dose. At least two components can be recognized: (a) variation in y due to the difference between the observed slope of the dosage-response curve and a slope of zero and (b) the variation about this fitted line. The second component can be determined directly. The "expected" Y is computed for each x from the equation of the line and subtracted from the corresponding observed y. The N differences are squared and summed to obtain $S(y - Y)^2$ as the sum of squares about the line. The N values of Y are computed with two statistics from the original N responses, $\bar{y}$ and b. Just as the mean $\bar{y}$ used up one degree of freedom in determining $[y^2]$, fitting the slope b as well restricts the $(y - Y)$'s to $N - 2$ degrees of freedom. The difference, $[y^2] - S(y - Y)^2$, with one degree of freedom represents that part of the total sum of squares which can be attributed to the slope of the dosage-response line.

Because of the additive property of the sums of squares, it is easiest to compute the effect of slope directly and obtain $S(y - Y)^2$ by difference. The missing term is easily derived from Eq. 1 by simplifying the right side of

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

which reduces to

$$S(Y^2) = \frac{S^2(y)}{N} + \frac{[xy]^2}{[x^2]}$$

where $[xy]$ and $[x^2]$ are defined in Eqs. 4 and 5. Each term on the right has one degree of freedom. The first is the correction for the mean in Eq. 8, so the second must represent that part of the variation in y accounted for by the slope of the dosage-response line. It will be designated as B^2 and defined as

$$B^2 = \frac{[xy]^2}{[x^2]} \tag{9}$$

With these terms, the work form for the simple analysis of variance in Table IV can be constructed. The original total about zero is given here as an additional line in the table but will be omitted in later analyses as of no intrinsic interest. The first column of the work form gives the source of variation from which each sum of squares is computed and

TABLE IV. Work Form for Analysis of Variance of the Variation about a Dosage-Response Line Computed from Individual Observations

Variation assigned to	D.F.	Sum of squares	Mean square	F
Slope of line, b	1	$B^2 = [xy]^2/[x^2]$ (Eq. 9)	B^2	B^2/s^2
Variation about line	$N - 2$	$[y^2] - B^2$	s^2	
Total about mean	$N - 1$	$[y^2] = S(y^2) - C$ (Eq. 8)		
Mean response $\bar{y}$	1	$C = S^2(y)/N$		
Original total about zero	N	$S(y^2)$		

the second column the degrees of freedom (D.F.) allotted to each component. Neither of these depends upon the results observed at the end of an experiment but rather upon its structure or design. Especially in tests of greater complexity it is good practice to construct this part of the table before starting an experiment as a check on the design, so as to insure that the basic terms can be isolated with an adequate number of degrees of freedom.

The sums of squares in the third column are computed from the response metameters. Two components represent complete equations, but most other work forms will give only the right side of the equation and usually without duplication by equations in the text. Each mean square

is computed by dividing the sum of squares in the same row by the degrees of freedom.

The first objective in Table IV is to determine the mean square (s^2) measuring the variation about the line, sometimes called the "error variance." The square root of s^2 or standard deviation is our best estimate from a given sample of the parameter σ which would characterize a similar dosage-response line with an infinite number of observations. In equation form,

$$s = \sqrt{\frac{[y^2] - B^2}{n}} \tag{10}$$

where n is the degrees of freedom, here equal to $N - 2$.

A second objective of Table IV is to determine whether the computed slope b differs significantly from no slope at all. This is equivalent to assuming that the true value of the slope is zero and that the observed value differs from it only by chance, the so-called null hypothesis. This null hypothesis is tested by computing the variance ratio. If there were no real relation between dose and response the mean square due to slope (B^2) should not exceed s^2 more than would be expected by chance. To determine its statistical significance, B^2 is divided by s^2 to obtain the variance ratio F in the last column. If the observed value of F is larger than 1, it is compared with the value expected a priori for $n_1 = 1$ (the degrees of freedom in the larger mean square) and $n_2 = N - 2$ (the degrees of freedom in the smaller mean square). The expected values for this comparison have been computed and tabled (1) for different levels of significance, including $P = .05$ for odds of 1 in 20, $P = .01$ for 1 in 100 and $P = .001$ for 1 in 1000, and for different combinations of n_1 and of n_2. The ratio of $F = B^2/s^2$ should be highly significant ($P \leqslant .01$) for an acceptable technique. Unless a known change in the dose of vitamin has had a real effect upon the response, the method is inadequate or the observations are too few for measuring unknown differences in potency. In practice any dosage-response curve which would be acceptable to the experimenter when plotted and examined visually would probably prove statistically significant by an analysis of variance, so that this phase of the calculation is less important than the determination of the error variance s^2.

Example 1b. The analysis of variance in Table IV may be illustrated numerically with the data of the dosage-response curve for vitamin K in Ex. 1a. From the y's in Table I and the statistics beneath it, we obtain in Table V the entries defined in Table IV. The standard deviation from the mean square in the second row of the analysis (or by Eq. 10),

TABLE V. Analysis of Variance of the Dosage-Response Curve for Vitamin K in Table I

Variation assigned to	D.F.	Sum of squares	Mean square	F
Slope of line, b	1	$-2.1170^2/1.1192 = 4.0044$	4.0044	429
Variation about line	13	.1214	.00934	1.00
Total about mean	14	4.1258		
Mean response $\overline{y}$	1	$24.36^2/15 = 39.5606$		
Original total about zero	15	43.6864		

is $s = \sqrt{.00934} = .0966$. The observed variance ratio for slope ($F = 429$) is so large as to leave no doubt as to its reality. In the absence of a real effect, an F as large as 17.81 would occur only once in 1000 experiments, as read from a table of the variance ratio at $P = .001$ for $n_1 = 1$ and $n_2 = 13$.

B. TESTS FOR HETEROGENEITY OF DOSE MEANS

In determining the dosage-response curve for a vitamin it is customary to record two or more responses at each of k doses. It is then possible to recognize two components in the variation about the fitted line. One is the scatter of the mean responses at successive dosage levels about the fitted line, and the other is the variation of the individual responses about these dose means. If the first component were to exceed the second significantly, the straight line could be considered an inadequate description of the relation between log dose and response. This contingency is tested by subdividing the sum of squares in the second line of Table IV into two components, one representing the deviations of the dose means from the line and the other the deviations of the individual response metameters about the dose means. The respective sums of squares are computed as shown in the work form in Table VI. If each

TABLE VI. Work Form for the Analysis of Variance of a Dosage-Response Curve Computed with Two or More Responses at Each Dose

Variation assigned to	D.F.	Sum of squares	Mean square	F
Slope of line, b	1	$B^2 = [xy]^2/[x^2]$	B^2	B^2/A
Scatter of dose means about line	$k - 2$	$S(T_d^2/f) - C - B^2$	A	A/s^2
Deviations about dose means	$N - k$	Remainder	s^2	
Total	$N - 1$	$[y^2] = S(y^2) - C$		
Correction for mean	1	$C = S^2(y)/N$		

dose total (T_d) is based upon the same number of observations, so that f is constant, f may be moved outside the parentheses and the second sum of squares computed as

$$\frac{S(T_d^2)}{f} - C - B^2$$

If f is constant and the doses are spaced at equal intervals on a log scale, so that the doses can be coded in computing the slope, B^2 is also calculated with the coded doses as

$$B^2 = \frac{S^2(x_1 T_d)}{f S(x_1^2)} \tag{9a}$$

where x_1 is the coded dose as defined on page 458. As before, each sum of squares is divided by its degrees of freedom to obtain the mean squares in the fourth column and in this case the mean square in the third row is our initial estimate of the random error variance s^2.

Additional information about the dosage-response curve is provided by the values of F in the last column of Table VI. As before, these are compared with the values which would be expected at a given probability if both variances in the ratio were random samples from the same population. If an observed F is larger than would be expected once in 20 similar trials, we conclude that it was computed from two variances which differed significantly. The test might show that the dose means scattered excessively about the straight line as judged from $F = A/s^2$. If so, we would conclude either that the dosage-response line was not linear or that the animals tested at the different dosage levels were not fully comparable. In this case A would be a more appropriate measure of the experimental error than s^2 for testing the significance of B^2. If A/s^2 should not exceed 1 significantly, both components might be included in the estimated standard deviation about the line as in Eq. 10.

Example 2b. The variation about the dosage-response curve in Table II for vitamin B_1 has been separated into two parts by the formulae in Table VI. Since the means varied less from the line on the average than

TABLE VII. Analysis of Variance for Dosage-Response Curve in Table II

Variation assigned to	D.F.	Sum of squares		Mean square
Slope of dosage-response line	1	$.8339^2/1.02217$	$= .6803$	.6803
Scatter of dose means about line	1	$49.3141/5 - 9.1801 - .6803$	$= .0024$	.0012
Deviations about dose means	2	$.7996 - .6803 - .0024$	$= .1169$	.00731
Total	16	$9.9797 - 9.1801$	$= .7996$	
Correction for mean	19	$13.55^2/20$	$= 9.1801$	
Composite error	18	$.0024 + .1169$	$= .1193$	$.00663 = s^2$

the individual measurements from the means, no F test was necessary to show that the curve for vitamin B_1 was effectively linear. We might prefer in this case to pool both components into a composite error as shown in the last row of the table, giving $s = \sqrt{.00663} = .0814$ with 18 degrees of freedom in accord with Eq. 10.

C. TEST FOR SIMPLE CURVATURE

If the mean square for the dose means exceeds that for the deviations about the dose means significantly, the question may arise as to how much of this larger mean square is due to erratic variation among the dose means and how much to systematic curvature. A convenient test for this purpose is to fit the parabola $Y = a' + bx + b'x^2$ to the data. If the three statistics of this equation a', b, and b' account for significantly more variation than the two statistics used in computing a straight line, curvature is present. Some observations may approach a ceiling or a floor, so that the curvature could be corrected by dropping a value at one or both ends. Alternatively, the relation may be smoothly curvilinear and require a change in the response metameter (or dose metameter) to convert it to a straight line. If the effect of b' or the "quadratic" term were not significant, this would not "prove" linearity, for with more observations the apparent curvature might become significant. However, if it should not approach significance, one may act as if the curve were a straight line.

If the parabola is computed in the form shown above, the calculation is impracticable in all but the most critical cases. However, when the doses are spaced evenly on the logarithmic scale and each dose is represented by the same number of observations, both x and x^2 can be replaced by the so-called orthogonal polynomials which give results proportional to b and b' but which are independent of each other. Moreover, the test can be extended to include higher powers of x. The procedure is especially useful in well-planned assays with three or more dosage levels, as will be shown later in this chapter. As applied to the analysis of variance of a single dosage-response curve, the sum of squares for the scatter of the dose means about the line is split into two parts.

The first step in the calculation is to replace the observed log doses x with the coded values x_1 described previously and the x^2's by similar coefficients x_2. We then compute the sum of the squares with one degree of freedom for quadratic curvature by the equation

$$Q^2 = \frac{S^2(x_2 T_d)}{fS(x_2^2)} \tag{11}$$

where each x_2 is the orthogonal polynomial ξ'_2 as given for a series of each specified length in Table XXIII of reference (1).

The sum of squares in the row for "scatter of dose means about the line" in Table VI is now divided into two components. The first is designated as "simple curvature" with one degree of freedom and its sum of squares is Q^2 as defined in Eq. 11. The second is labeled "scatter of dose means about the curve" with $k - 3$ degrees of freedom, and it is the residual variation from the second row of Table VI, so that it is computed as

$$\frac{S(T_d^2)}{f} - C - B^2 - Q^2$$

If Q^2 does not exceed significantly or near significantly either the residual scatter of the dose means about the simple curve or s^2, the deviations about the dose means, one may prefer to segregate the quadratic component only in preliminary working tables.

Example 3b. The dosage-response curve for niacin on page 459 meets the formal requirements for a test of simple curvature and the quadratic coefficients x_2 for a series of 9 are shown in the last column of Table III. These values are orthogonal with the coded doses x_1 in that $S(x_2) = 0$ and the sum of the products of $S(x_1x_2) = 0$. The computation of the analysis of variance from these data is shown in full in Table VIII. The mean squares in the second and third rows of the table have

TABLE VIII. Analysis of Variance of Dosage-Response Curve for Niacin in Table III

Variation assigned to	D. F.	Sum of squares		Mean square	F
Slope of straight line, b	1	$24.08^2/(3 \times 60)$	= 3.22137	3.22137	
Quadratic curvature	1	$1.12^2/(3 \times 2772)$	= .00015	.00015	.48
Scatter of dose means about parabola	6	153.9932/3 — 48.10673 — 3.22137 — .00015	= .00282	.00047	1.52
Deviations about dose means	18	3.22987 — 3.22137 — .00015 — .00282	= .00553	.00031	1.00
Total	26	51.3366 — 48.10673	= 3.22987		
Correction for mean	1	$36.04^2/27$	= 48.10673		
Composite error	25	.00015 + .00282 + .00553	= .00850	.00034	

been divided by the deviation about the dose means in the last row. The largest variance ratio, $F = 1.52$, would be expected oftener than once in five trials on the null hypothesis, so that there is no evidence of heterogeneity or curvature in the present series and we would probably pool all variation about the line in computing $s^2 = .00034$. The variance at-

tributable to the slope (B^2) was here so very much larger than s^2 that a test of its significance would be superfluous.

3. The Precision of the Line

After the dosage-response line has been computed from the individual responses and their variability about the line has been tested for homogeneity and measured, we are ready to evaluate the method as an assay technique. The precision of the line calculated from a given set of data depends not only upon the variation about the line and its slope but also upon the number of observations and the way the doses are distributed in the experiment. All four factors enter in the standard error of the line. If the dosage-response curve is to form the basis of an assay, however, we will be concerned not in predicting a future response from a known dose but rather in using the observed response to predict an unknown dose. To measure the inherent precision of inverted estimates of this type, the standard deviation s of a single response, as measured initially in units of y, must be converted into units of x or log dose, in which terms it will be designated as λ. Finally, we need to consider how λ can be used in designing assays.

A. THE STANDARD ERRORS OF THE COMPUTED LINE

An analysis of variance has shown, let us suppose, that a given dosage-response curve can be considered as linear within a specified range of doses and that the variation about the line can be treated as homogeneous and estimated by s^2. The validated straight line has been computed from a specific sample of observations which theoretically could be increased indefinitely in number. If conditions were stable, the aggregate of an infinite number of such samples would represent a "population." Our interest in a particular sample depends upon how good a representative it is of the "population" from which it was drawn. Similarly, we can look upon the statistics a and b which define the line in a given experiment as estimates of the parameters α and β in the population from which the sample was drawn. Although the parameters are unknown we can compute limits from a single sample which in a given proportion of all similar experiments would be expected to enclose the true or population values of a and b.

The limits depend upon the so-called standard errors of a and b, which, in turn, depend upon the total number of observations N and the sum of squares of the log doses $[x^2]$. The standard error of the position of the line as measured by a is

$$s_a = \sqrt{s^2/N} \tag{12}$$

The standard error of the slope b is

$$s_b = \sqrt{s^2/[x^2]} \tag{13}$$

Since a and b are presumably uncorrelated, the value of Y predicted from Eq. 1 at any contemplated X is subject to variability easily predicted from both standard errors, so that

$$s_Y = \sqrt{s_a^2 + s_b^2 (X - \bar{x})^2} \tag{14}$$

The more dose X differs from the observed mean $\bar{x}$, the greater is the influence of the error in the slope upon the Y predicted from the equation of the line.

The interpretation of the standard error in each of the above cases is the same. It determines a range above and below each statistic which in about two out of three experiments will bracket the unknown true value or parameter in the population from which the sample has been drawn. The range for other odds can be computed by multiplying the standard error by a coefficient which depends upon the required level of significance P and the degrees of freedom n used in estimating the standard deviation s. This coefficient is known as Student's t. It has been tabled by Fisher (1, 3) in a convenient form for different combinations of P and n. When the standard error for a, b, or Y is multiplied by t and the product is added to and subtracted from the corresponding statistic, the resulting values are known as confidence or fiducial limits. In applications other than those considered in this chapter, fiducial and confidence limits are not always identical and they differ considerably in theory. Here they are arithmetically the same.

These limits are in units of the response y. They can be used, for example, to compute limits Y_L enclosing the true response as predicted for different values of X by the equation for the dosage-response curve,

$$Y_L = Y \pm ts_Y \tag{15}$$

where Y is the expected response as computed by Eq. 1 and s_Y is defined by Eq. 14. t is read from Table III in reference (1) or its equivalent at the desired probability P of *not* bracketing the true value with n = the degrees of freedom in s^2. As we shall see later, only approximate limits can be obtained in this manner in units of the log dose x. The exact limits in such cases are described in later sections.

Example 2c. The dosage-response curve for vitamin B_1 in Table II provides a convenient illustration of the standard errors of a computed line and their use in computing confidence or fiducial limits. The analysis of variance in Table VII has led to an estimated variance about the line of $s^2 = .00663$ with 18 degrees of freedom. The line has been

determined from 20 observations with a slope for which $[x^2] = 1.0222$. Given these terms, the standard errors of the line may be computed as

$$s_a = \sqrt{.00663/20} = \sqrt{.0003315} = .0182 \quad \text{by Eq. 12,}$$
$$s_b = \sqrt{.00663/1.0222} = \sqrt{.006486} = .0805 \quad \text{by Eq. 13}$$
$$\text{and} \quad s_Y = \sqrt{.0003315 + .006486(X - .3450)^2} \quad \text{by Eq. 14.}$$

It is of interest to draw confidence limits above and below the computed line which will define a band enclosing the expected response. These are computed by Eq. 15 as shown in Table IX. The shortened equation of the line (Table II) is $Y = .3960 + .8158X$ and $t = 2.878$ for $n = 18$ and $P = .01$.

TABLE IX. Calculation of Confidence Limits for Dosage-Response Curve for Vitamin B_1 in Table II

Log dose (−1)	Expected response	Steps in solving Eq. 14				Confidence limits at $P = .01$	
X	Y	$X - \bar{x}$	$(X - \bar{x})^2$	s_Y^2	s_Y	$Y + ts_Y$	$Y - ts_Y$
−.1	.314	−.445	.1980	.001616	.0402	.430	.198
.0	.396	−.345	.1190	.001103	.0332	.492	.300
.1	.478	−.245	.0600	.000721	.0269	.555	.401
.2	.559	−.145	.0210	.000468	.0216	.621	.497
.3	.641	−.045	.0020	.000344	.0185	.694	.588
.4	.722	.055	.0030	.000351	.0187	.776	.668
.5	.804	.155	.0240	.000487	.0221	.868	.740
.6	.885	.255	.0650	.000753	.0274	.964	.806
.7	.967	.355	.1260	.001149	.0339	1.065	.869

The upper and lower confidence or fiducial limits in the last two columns of Table IX have been plotted against X in Fig. 1 and connected by broken lines. They define two parabolae which diverge least from the computed line at the point $\bar{x}$, $\bar{y}$. Limits calculated in this way will enclose any one point on the true dosage-response curve in 99 out of 100 experiments. In contrast, the individual observations should and do define a broader zone of uniform width above and below the fitted line.

B. THE STANDARD DEVIATION IN TERMS OF THE LOG DOSE, λ

In determining the potency of a vitamin, the dosage-response curve is used in reverse and X is estimated from an observed y. For this reason it might be argued that we should determine the slope of the line relating X to Y instead of that relating Y to X. This, however, would be a mistake. The form of the dosage-response curve is fixed by the logic of the experiment. The response is the dependent variate and is subject to a sampling error reflecting the varying susceptibility of the test

animals. The dosage of vitamin on the other hand is the independent variate with values determined arbitrarily and more or less accurately by the experimenter. Hence the curve is computed with the response as a function of the log dose even though it may be used to estimate the dose corresponding to a selected or observed response. For this purpose Eq. 1 is rewritten as

$$X = \bar{x} + \frac{Y - \bar{y}}{b} \tag{1c}$$

The standard deviation in X depends upon the standard deviation in Y divided by the slope, or, following Gaddum's terminology, upon

$$\lambda = \frac{s}{b} \tag{16}$$

The magnitude of λ determines the efficiency of a given response as a method for assaying vitamin potency. Hence a modification in procedure which either reduces the standard deviation s or increases the slope b should decrease the assay error. As noted in the first part of the chapter on animal assays, λ has been used throughout the discussion of the animal assays as the major criterion for comparing alternative methods. The smaller the value of λ, the greater is the inherent precision of an assay technique.

Although represented by a Greek letter, λ is a statistic computed from a sample and as such has a standard error based upon the standard errors of its constituents. The standard error of b is given by Eq. 13; that of s is equal approximately to

$$s_s = \sqrt{s^2/(2n + 0.5)} \tag{17}$$

where n is the number of degrees of freedom in s^2. On the assumption that s and b are not correlated, the standard error of their ratio may be computed as

$$s_\lambda = \lambda \sqrt{\frac{1}{2n + 0.5} + \frac{s^2}{B^2 - s^2 t^2}} \tag{18}$$

where t is the value expected at a given P, n is the degrees of freedom in s^2, and B^2 is defined in Eq. 9. Both in determining λ and in the biological assay of a vitamin it is especially important that the slope of the dosage-response curve differ significantly from zero, preferably at $P \leqslant .01$. For this reason, an experiment in which the last term in Eq. 18 is negative, even when computed with t for $P = .05$, is inadequate for estimating λ.

Example 2d. The precision of the bradycardia technique for vitamin

B_1 may be estimated by Eq. 15. From Table VII, $s = \sqrt{.00663} = .0814$ and from Table II, $b = .8158$, so that $\lambda = .0814/.8158 = .0998$. The standard error of λ by Eq. 18 is computed as

$$s_\lambda = .0998 \sqrt{\frac{1}{36.5} + \frac{.00663}{.6803 - .00663 \times 4.414}} = 0.0194$$

where $t^2 = 4.414$ at $P = .05$. In this case the slope is so highly significant that increasing the odds for t^2 has little effect upon s_λ. At $P = .001$, for example, $t^2 = 15.382$ and $s_\lambda = 0.0197$.

C. THE USE OF λ IN PLANNING ASSAYS

A useful application of the dosage-response curve is in estimating the precision of an assayed potency based on a given number of observations or, conversely, in determining the number of observations needed to assay an unknown preparation with a given precision. Both of these estimates depend upon setting up the assay with an assumed potency of the unknown that is approximately correct. To the extent that this assumption proves incorrect, the number of observations computed from the equation will be too small. We will further assume that parallel tests are conducted with the unknown sample and a reference standard and that the experimental material is assigned equally to the standard and to the unknown.

The precision of an assay is a function of s_M, the standard error of the log ratio of potencies, M. The calculation of s_M from actual assays will be considered later. Under the above limitations, it can be estimated approximately for a projected assay based upon a total of N observations as

$$\text{Approximate } s_M = \frac{2\lambda}{\sqrt{N}} \tag{19}$$

where λ is determined with Eq. 16.

The standard error of a statistic is used to determine a range of values which may be expected to include the parameter estimated by a statistic with a given frequency. Thus the true log ratio of potencies would be expected to fall within the interval from $M - s_M$ to $M + s_M$ in 68 determinations in 100 and within the interval $M \pm 2s_M$ in about 95 determinations in 100. Although these limits are readily converted to units of potency by taking their antilogarithms, it is often convenient instead to multiply and divide a contemplated potency by the antilogarithm of s_M or of $2s_M$ to obtain the same limits. Alternatively, one can express the standard error of potency in percentage terms by com-

puting 100(antilog s_M − 1) and this practice has been followed frequently in the present book.

An estimate of the total number of observations (N) needed for a given precision may be computed as

$$N = \frac{4\lambda^2}{s_M{}^2} \tag{20}$$

In this case s_M is an assumed value which is determined by the precision that the experimenter requires. If, for example, an average standard error of ± 10% is acceptable, s_M would be the logarithm of 1.10 or 0.0414. The number of observations computed by Eq. 20 is as subject to sampling error as λ upon which it depends. To insure against underestimating the number of observations needed for a given precision in an isolated or critical assay, it is well to allow for the standard error of λ and increase the number of observations accordingly. One way in which this can be done is to compute N as

$$N = \frac{4(\lambda + ts_\lambda)^2}{s_M{}^2} \tag{20a}$$

where t is selected at a specified margin of safety. In the long run this equation will tend to overestimate the number of observations needed, but in critical cases it may insure against planning too small an experiment.

Example 2e. The expected precision of an assay based on the dosage-response curve for vitamin B_1 in Table II can be computed with λ = .0998 and N = 20. From Eq. 19, approximate $s_M = 2 \times .0998/\sqrt{20}$ = .0446, indicating a percentage standard error of 100(1.108 − 1) = 10.8%. For an assay with an average standard error of 10%, $N = 4 \times .0998^2/.0414^2 = 23.2$ or 24 observations. Since s_λ is here about 20% of λ, one might prefer to increase the number of observations to guard against a possible underestimate of λ by using Eq. 20a with t = 2.101 at P = .05. In this case $N = 4(.0998 + 2.101 \times .0194)^2/.0414^2 = 46$ or nearly twice as many responses as before.

III. Designs for Segregating Nonrandom Variation

Vitamin assays are subject to variation from many sources. Some of these can be controlled experimentally, and the success of an assay technique depends in large part upon their elimination. Other sources of variation can be controlled statistically by the adoption of a suitable design. By identifying the latter variables and segregating their effect, the experimental error can often be reduced materially. The types of

variation segregated in a good experimental design are usually qualitative, although occasionally quantitative factors may be balanced similarly. A technique cannot be considered as fully developed until the only variation affecting the results is the error of random sampling. There are many designs for this purpose but only two of them will be considered at this point. For an excellent discussion of the principles involved, Fisher's "The Design of Experiments" (19) is recommended.

1. Randomized Groups

One of the simplest and most effective experimental designs is that known as randomized groups or randomized blocks. It is applicable when the experimental material can be sorted in advance of testing into relatively homogeneous groups. Litter mates of the same sex, for example, may react more nearly alike to a vitamin than animals of different litters and sexes. This is recognized in the U.S.P. assay for vitamin A which requires a balancing of litter mates and of males and females between treatments. In other assays, such as the curative tests with vitamin B_1, an individual test animal can be used repeatedly. In this case groups might consist of the successive responses of single individuals.

However the group may be formed, two principles are followed. The first is to use equal-sized groups with as many individuals in each group as there are different doses of vitamin under test. In each group one individual is assigned to every treatment. Hence each group supplies a complete replicate, and precision is obtained by replicating complete groups. The differences between groups do not bias the mean response to each treatment, since each group is represented equally at all dosage levels. Mean differences between groups are also eliminated from the estimate of the experimental error, most readily by the analysis of variance. The first sum of squares then measures the variation due to differences between groups and the other sums of squares parallel those in the work form of Table VI for the dosage-response curve of a single preparation of vitamin. The principle is equally applicable to comparisons of a standard and an unknown as will be shown in the next section. If the groups really differ in their response, the mean square for the first term should be significantly larger than the experimental error in the last term of the analysis.

The second principle is to assign treatments at random to the individuals in each group. It is *not* sufficient to "think up" a random arrangement; an objective physical process is required. The experimenter may shuffle playing cards and, as the animals are picked, assign

the treatment represented by the successive cards as they are turned up. He may throw dice instead or use a table of random numbers. Without some such precaution one is in danger of unconsciously biasing the results as, for example, by giving the weakest animal of the litter the largest dose of vitamin and vice versa. The only restriction is that each treatment occurs only once within a group and that every group includes all treatments.

The analysis of variance of an experiment with f replicates or groups and k responses in each group may be computed by the work form in Table X. The sum of squares between groups is determined from the group totals T_g of the k responses in each of the f groups. The sum of squares for doses has been separated here into two components, although, if desired, the effect of simple curvature as defined in Eq. 11 (Q^2) can often be isolated from the scatter of the dose means. The error term is

TABLE X. Work Form for an Analysis of Variance of a Dosage-Response Curve in Randomized Groups

Term	D.F.	Sum of squares	Mean square	F
Differences between group totals	$f-1$	$S(T_g^2)/k - C$	G	G/s^2
Slope of the dosage-response line	1	$[xy]^2/[x^2]$	B^2	B^2/s^2 or B^2/A
Scatter of the dose means about the line	$k-2$	$S(T_d^2)/f - C - B^2$	A	A/s^2
Interaction of groups $\times$ doses	$(f-1)(k-1)$	Remainder	s^2	
Total	$N-1$	$[y^2] = S(y^2) - C$		
Correction for mean	1	$C = S^2(y)/N$		

the "interaction" of groups $\times$ doses. It includes both the variation in the slope of the dosage-response curve from group to group and the scatter of the separate observations about the straight lines for each group. These components are also easily separable but usually are combined as in Table X. The interaction mean square is used to test the adequacy of the straight line and to measure the advantage gained by the subdivision into groups. It is the error variance, s^2, required in the equations of the preceding sections.

Example 4a. The randomized group design is exemplified by a dosage-response curve for the growth assay of vitamin A in female rats. The response metameter in Table XI is the average growth per week in grams during a four-week assay period. Each one is the slope of a straight line fitted with the coefficients $x_1 = -2, -1, 0, 1$, and 2 to the five weekly

weighings from the beginning to the end of the assay period. Thus the rat at the smallest dose in the first litter weighed 105, 110, 109, 114, and 113 g. at the start of the test period and at the end of the first to fourth weeks respectively. These weights were multiplied in turn by the successive values of x_1 and the products summed to obtain $-(2 \times 105) - 110 + 0 + 114 + (2 \times 113) = 20$. Dividing by $S(x_1^2) = 10$ gives $20/10 = 2.0$ g. per week for the first entry in the table. The other entries were computed similarly.

TABLE XI. A Dosage-Response Curve for Vitamin A in Randomized groups, in terms of the Rate of Growth of Female Rats during a Four-Week Test Period

Data from the Squibb Laboratories

Dose mg./day	x_1	Growth, g./week, in litter No. 1	2	3	4	5	6	7	8	9	10	11	12	T_d	Assay dosage
1.06	−3	2.0	1.4	3.1[a]	−1.2	−8.6	5.9	3.5	−8.6	1.1	4.7	0.6	4.2	8.1	U_1
1.50	−1	0.8	5.4	5.9	5.5	0.8	1.3	3.8	2.1	6.4	10.7	3.0	8.8	54.5	S_1
2.12	1	1.4	5.3	9.3	−3.1	6.8	3.6	3.0	2.5	12.7	8.7	9.4	2.1	61.7	U_2
3.00	3	6.0	13.6	11.2	4.1	5.4	11.2	10.3	8.0	12.5	11.8	9.5	10.3	113.9	S_2
T_g		10 2	25.7	29.5	5.3	4.4	22.0	20.6	4.0	32.7	35.9	22.5	25.4	238.2	

[a] Missing value, see p 481. Assumed assay dosages in last column are used in example 4c.

$\bar{x} = (\log 1.50 + \log 2.12)/2 = .2512$; $\bar{y} = 238\ 2/48 = 4.9625$

$S(x_1T_d) = 324\ 6$; $fS(x_1^2) = 12 \times 20 = 240$; $I' = .1505/2$

$b = 2 \times 324\ 6/(.1505 \times 240) = 17.97$ (Eq. 7)

The four dosage levels were repeated in all 12 litters, so that the total response for each group or litter (T_g) represented the same combination of treatments. Similarly, every litter was represented equally in the four dose totals (T_d), so that litter differences could not bias the response to dose of vitamin. The largest dose of 3 mg. daily of U.S.P. reference standard was divided successively by $\sqrt{2}$ to obtain the three smaller doses in a geometric series.

The analysis of variance in Table XII has been computed as shown in the work form of Table X with one exception. The degrees of freedom in the total and in the error have each been diminished by 1 to adjust for the missing rat in litter No. 3, which has had to be computed

TABLE XII. Analysis of Variance of the Data on Vitamin A in Table XI with the Work Form in Table X

	D.F.	Sum of squares	Mean square	F
Between litters	11	346.96	31.54	3.13
Slope	1	439.02	439.02	43.51
Scatter of dose means	2	30.24	15.12	1.50
Error	32	323.01	10.09	1.00
Total	46	1139.23		
Correction for mean	1	1182.07		

from the remaining data as described later (p. 481). The mean square between litters is three times as large as the interaction of litters × doses. If rats had been assigned to treatments at random without respect to litter differences, this variability would have been included in an enlarged experimental error and the number of animals would have had to be increased by 50% to obtain the same precision. Other experiments on vitamin A (p. 77) have shown a similar relation between and within litters, fully justifying the segregation of litter differences. The effect of slope was computed from Eq. 9a as $B^2 = (324.6)^2/240 = 439.02$ and the scatter of dose means by difference from the sum of squares for dose as $(8.1^2 + 54.5^2 + \cdots + 113.9^2)/12 - 1182.07 - 439.02 = 30.24$. It is evident from the variance ratios in the second and third rows that the dosage-response curve could be fitted adequately by a straight line, which has the equation $Y = 4.96 + 17.97\ (X - .251)$.

2. The Latin Square

Two potential sources of variation may occur independently in an assay technique. Thus if the test animal can be used more than once, as in the vitamin B_1 curative assay, the response of different individuals may vary much more than would be expected from repeated tests on the same individual. At the same time the animal's sensitivity may change during the assay, so that successive responses tend to increase or decrease progressively. Under these and similar circumstances we may wish to segregate both the variation between individuals and that due to order of treatment. This can be done by a well-known experimental design called the Latin square.

In the Latin square there are two restrictions upon the randomization of treatments. These restrictions are represented schematically by the rows and columns of a table in which the unit square contains an equal number of rows, columns and letters. Each letter represents one treatment and occurs once in every row and in every column. In the biological assay of vitamins, the 4 x 4, 5 x 5 and 6 x 6 Latin squares are the most useful. The typical squares in Table XIII have been taken from the basic Latin squares in Table XV of reference (1).

Table XIII. Typical Latin Squares before Randomization

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

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

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

Each must be randomized before it is used and by this process each will generate many different squares. The order of rows is first randomized, then that of the columns and finally the letters are assigned at random to the several doses. These steps may be illustrated for a 4 x 4 square as follows:

A	B	C	D			A	B	C	D			B	C	A	D
B	A	D	C	Rearranging → rows		C	D	A	B	Rearranging → columns		D	A	C	B
C	D	A	B			D	C	B	A			C	B	D	A
D	C	B	A			B	A	D	C			A	D	B	C

Latin squares should be randomized independently each time one is used.

In vitamin research a single Latin square may not give the necessary precision. Several Latin squares can be used simultaneously, however, so that the designation for each row extends over all squares while the columns differ. Thus if the columns represent different individual animals and rows the order of treatment, we might have n' Latin squares, each with k rows, columns, and treatments. Since the restriction represented by rows would be common to all squares, there would be k rows, k treatments, and $n'k$ columns. The analysis of variance for a dosage-response curve then takes the form in Table XIV, where T_c, T_r, and T

TABLE XIV. Work Form for Analysis of Variance of a Dosage-Response Curve in n' Latin Squares

Term	D.F.	Sum of squares	Mean square
Between columns	$n'k - 1$	$S(T_c^2)/k - C$	
Between rows	$k - 1$	$S(T_r^2)/n'k - C$	
Slope of dosage-response line	1	$[xy]^2/[x^2]$	B^2
Scatter of dose means about the line	$k - 2$	$S(T_d^2)/n'k - C - B^2$	A
Remainder or error	$(n'k - 2) \times (k - 1)$	By difference	s^2
Total	$N - 1$	$S(y^2) - C$	
Correction for mean	1	$T^2/N = C$	

symbolize column totals, row totals, and the grand total respectively. When $n' = 1$, this reduces to the analysis for a single Latin square.

When treatments are assigned by means of a Latin square, an unbiased estimate of the error can be assured only if the sum of squares for each restriction in the design is segregated from the remainder or error. In cases where one restriction represents a graded variate, such as increasing body weight, the experimenter may prefer to use covariance to isolate and measure its effect, as will be described presently. When

official assays require the balancing of two factors, such as litters and body weight at depletion, Latin squares provide a convenient method for assigning animals to treatments. If the mean square for one of these restrictions, body weight for example, should prove repeatedly of the same magnitude as the remainder, the experimenter might retain the Latin square design as a device for meeting official requirements but ignore this restriction in computing the assay error. This could lead to a small systematic overestimate of the error which might be preferred in routine assays to the additional calculation needed for segregating the second restriction in the analysis.

Example 5a. The data of a dosage-response curve for vitamin B_1 from the length of cure of polyneuritis are given in the two 4 x 4 Latin squares in Table XV. The doses corresponding to the treatments in the left side of the table are shown at the right; the columns represent individual rats and rows the order of treatment. Rats 5 and 8 died before the fourth test was completed, so that these values have been replaced

TABLE XV. Dosage-Response Curve for an Assay of Vitamin B_1 by the U.S.P. Method—Response of rats Measured in Length of Cure of Polyneuritis Data from Food Research Laboratories

Order of	Dose in rat No								Days cured, rat No.								Dosage				
																		μg. of			
treatment	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	T_r	B_1	Symbol	x_1	T_d
I	A	D	C	B	A	D	C	B	15	8	3	2	17	13	11	2	71	3 00	*B*	—3	23
II	D	A	B	C	B	C	D	A	11	13	3	7	5	4	11	12	66	4 24	*C*	—1	59
III	C	B	A	D	D	A	B	C	12	1	13	10	11	12	5	7	71	6 00	*D*	1	78
IV	B	C	D	A	C	B	A	D	2	6	6	10	9[a]	3	19	8[a]	63	8 48	*A*	3	111
T_c									40	28	25	29	42	32	46	29	271 = *T*				271

[a] Missing values, replaced as described on p. 481.

by successive application of Eq. 22 as described in the next section. In consequence, the degrees of freedom in the error and in the total have been diminished by 2. The analysis of variance in Table XVI followed the work Form of Table XIV.

TABLE XVI. Analysis of Variance of the Data on Vitamin B_1 in Table XV with the Work Form in Table XIV

	D.F.	Sum of squares	Mean square	*F*
Order of treatment	3	5.85	1.95	.44
Between rats	7	103.72	14.82	3.32
Linear	1	500.56	500.56	111.98
Curvature	2	6.28	3.14	.70
Error	16	71.56	4.47	1.00
Total	29	687.97		
Correction for mean	1	2295.03		

We would conclude from the analysis that the relationship between length of cure in days and the log dose of vitamin B_1 was satisfactorily linear. Order of treatment apparently did not modify the response, since its mean square was less than that for the error, but the differences between rats exceeded the error more than threefold. If an assay were to be based upon the present results, one might prefer randomized groups with each group representing the four responses of a single rat. This illustrates how Latin squares can be used to test simultaneously in a single experiment the importance of two suspected sources of variation.

3. The Replacement of Missing Values

The results of a balanced experiment are sometimes incomplete due to the loss of one or more animals. This occurred in both of the two preceding examples. Such losses destroy the balance upon which the analysis of an experiment depends for its simplicity. Fortunately, the balance can be restored by replacing each missing value with a number y' computed from the rest of the data. The number is computed so as to minimize the sum of squares for error in the completed analysis. The general equation is given in reference (19); the discussion here will be restricted to the designs described above.

In an experiment in randomized groups the replacement for a single missing value is computed as

$$y' = \frac{kT_d' + fT_g' - T'}{(k-1)(f-1)} \tag{21}$$

where there are k doses, f groups, and T_d', T_g', and T' are the totals of the response metameter for the dose with the missing value, the group with the missing value, and the total of all known values respectively. The replacement is entered in the empty place in the table and the totals and the sums of squares are computed just as if there had been no loss. The sum of squares for error, however, loses one degree of freedom for every value which must be computed and with the reduction in the degrees of freedom, the mean square for error is an unbiased estimate of s^2. The substitution results in an overestimate in the sum of squares for treatments but with one or few replacements the increase is small and is usually neglected. Methods for correcting this bias are given in reference (20).

Incomplete Latin squares are analyzed by a similar device. Each missing value is replaced by a number computed as

$$y' = \frac{k(n'T_c' + T_r' + T_d') - 2T'}{(k-1)(n'k-2)} \tag{22}$$

where T_c', T_r' and T_d' represent the totals for columns, rows, and doses with the missing value. The computed estimate is used just as in the case of randomized groups.

When an experiment in randomized groups or Latin squares has two or more missing observations, the replacements are computed most readily by successive approximations. The mean of the observed values at the same dose is inserted temporarily in each gap except one, and the missing value for the remaining space is then computed with Eq. 21 or 22. Next, one of the arbitrary numbers is erased and replaced by the same equation, continuing the process until each has been computed. This first set is a first approximation, and the cycle is repeated until successive repetitions lead to no change in the computed values. Every missing number, however, reduces the degrees of freedom in the error by 1 and enlarges the bias in the other sums of squares.

Example 4b. The missing value in litter 3 of Table XI was computed by Eq. 21 as

$$y' = \frac{4 \times 5.0 + 12 \times 26.4 - 235.1}{(4-1)(12-1)} = 3.1$$

As a result the error term in the analysis of variance in Table XII had 32 instead of 33 and the total 46 instead of 47 degrees of freedom.

Example 5b. Two values were missing from the Latin squares in Table XV, both on the fourth test with rats 5 and 8. The seven observed responses for dose C had an average of $50/7 = 7$, which was substituted temporarily in computing the first approximation for the response of rat 8 to dose D by Eq. 22 as

$$y' = \frac{4(2 \times 21 + 53 + 70) - 2 \times 261}{(4-1)(2 \times 4 - 2)} = 8$$

With this computed value in the gap for rat 8 the missing response for rat 5 was calculated next as $y' = 9$. Recalculation of the response for rat 8 did not change the initial estimate, so that the totals in Table XV could be completed with these values. The two replacements reduced the degrees of freedom in the rows for error and total in Table XVI from 18 to 16 and from 31 to 29 respectively.

IV. Measurement of Relative Potency

The dosage-response curve for a standard preparation of vitamin is determined primarily as a guide in developing a satisfactory experimental technique and for selecting a suitable dosage range. The response to known doses of vitamin is so variable that such a curve cannot

be used directly to estimate the potency of unknown preparations. For quantitative results both the unknown and the standard are tested in every assay. Moreover, one or both preparations should be tested at two or more dosage levels since the slope of the dosage-response curve may vary not only from one laboratory to another but also from time to time in the same laboratory (21). In consequence, an assay should provide estimates of (a) the mean response to corresponding doses of standard and unknown, (b) the slope of the dosage-response curve, and (c) the experimental error s^2. When the slope and the experimental error show good stability over a period of time or between different laboratories, this information may be utilized to increase the reliability of a single assay, as shown in a later section. Even in this case, however, it is good insurance to design each individual assay so that it is self-contained.

An initial study of the dosage-response curve should cover a relatively wide range of doses in order to define the zone in which the response is a linear function of the log dose. Assays falling within a known range need not include as many dosage levels as the initial experiments. Before setting up an assay, the anticipated potency of the unknown is estimated from other sources, and this assumed potency forms the basis for administering the standard and the test preparation or unknown.

1. Determinations of Potency without Restrictions in Dose

Relative potency can be computed from assays with few or no refinements in design, although in most cases the factorial assays described in the next section are preferred. The more general types may be considered under two headings, comparisons of each unknown at a single dosage level with a concomitant standard curve based upon two or more doses and comparisons of the dosage-response curves of two preparations.

A. Assays with one dosage level of each unknown

It may be necessary to determine the potency of a vitamin preparation of which very little is available, or to test the activity of several different fractions in purifying a compound. A regulatory laboratory may have many samples to assay for agreement with claim. In these and other cases each unknown may be administered by necessity or by choice at only one dosage level and the standard at several dosage levels in the same experiment. If the response to an unknown is in the range covered by the linear part of the dosage-response curve of the standard, its potency can be determined readily.

The potency computed from such an assay, however, may be of limited validity. It represents a single level of response and there is no

evidence that the potency would be the same, within limits of the sampling error, at larger or smaller doses. In extending his result to other dosage levels, the experimenter assumes that, if determined, the slope of the dosage-response curve for the unknown would have been the same as that of the standard. In other words, he assumes that under his assay conditions the unknown differs from the standard *only* in its concentration of the same active constituent. When both standard and unknown are given at two or more dosage levels, this assumption can be tested as part of the assay.

With one dose of the unknown, the responses are often distributed less efficiently than in the factorial assays. The same number of animals is usually assigned to the unknown and to each dose of the standard, which weights them quite unequally. In assaying a single unknown, the most efficient distribution is to divide the animals equally between the standard and the unknown. However, if there are g unknowns in the same assay, each tested with N responses, the total number of responses at all dosage levels of the standard should be $N\sqrt{g}$ for the greatest precision over the entire assay (22). Thus in an assay with one unknown ($g = 1$), N responses would be assigned to both the unknown and the standard; in an assay with four unknowns ($g = 4$), N responses would be assigned to each unknown and $2N$ responses to the standard.

For a single dosage level of the unknown, potency is determined from the dosage-response curve of the standard and the mean response of each unknown. The dosage-response curve is computed as described earlier in this chapter, leading to $\bar{x}_s$ the mean log dose, and $\bar{y}_s$ the mean response at all dosage levels, and b_s the slope of the line, the subscript s designating "standard." With these terms and the mean response for a given unknown $\bar{y}_u$ at the single log dose level x_u, the log ratio of potency M may be computed as

$$M = \bar{x}_s - x_u - \frac{\bar{y}_s - \bar{y}_u}{b_s} = \bar{x}_s - x_u + \frac{\bar{y}_u - \bar{y}_s}{b_s} \tag{23}$$

The antilogarithm of M is the required potency of the unknown.

Example 6a. The A.O.A.C. chick assay for vitamin D (p. 126) requires testing the reference standard at three or more dosage levels, but each unknown at only one level Usually many feeding oils are tested against the standard in the same assay, and Table XVII gives the results for six unknowns and four levels of the standard in a recent routine assay at the Connecticut Agricultural Experiment Station. The chicks were reared in two brooder batteries, each of 16 cages, with nine chicks in

TABLE XVII. A Vitamin D Assay by the Chick A.O.A.C. Method with Six Unknowns Each at a Dosage Level of 9 Units. Data from R. B. Hubbell, Connecticut Agricultural Experiment Station.

Preparation of vitamin D	Units[a] per 100 g. of ration	% Tibia ash (— 30) in battery		Total	Mean ash			Potency 100 (anti-log M)
		I	II	T_d	$\bar{y}$	$\bar{y}_u - \bar{y}_s$	M	
None	0	4.2	4.1	8.3				
Reference Standard	4	8.6	6.8	15.4				
	6	11.5	10.4	21.9	10.75			
	9	14.6	12.6	27.2				
	13.5	15.0	15.2	30.2				
Unknown A	9	11.4	12.1	23.5	11.75	1.00	— .116	77
" B	9	7.3	6.6	13.9	6.95	— 3.80	— .403	40
" C	9	12.5	12.4	24.9	12.45	1.70	— .075	84
" D	9	14.8	13.6	28.4	14.20	3.45	.030	107
" E	9	9.7	9.8	19.5	9.75	— 1.00	— .236	58
" F	9	14.3	10.8	25.1	12.55	1.80	— .069	85
Total (T_g,T)		123.9	114.4	238.3				

[a] Unitage of unknowns based upon claimed content of vitamin D in each feeding oil.

$$S(x_1T_d) = -15.4 + 27.2 = 11.8;\ i = .176,\ b_s = 11\,8/(.176 \times 2 \times 2) = 16.75$$
$$S(x_2T_d) = +15.4 - 2 \times 21.9 + 27.2 = -1.2;\ Q^2 = (-1.2)^2/(6 \times 2) = .12$$
$$\bar{x}_s = \log 6 = .778,\ \bar{x}_u = \log 9 = .954$$

most cages and only three cages varying by more than one chick from this number. For reasons described on p. 128, small variations in the number of chicks per cage may be ignored and the response for each cage weighted equally in computations of potency. Each battery provided one complete replicate with the treatments assigned independently and at random to the cages in each battery.

The response to the standard in terms of the average per cent ash in the pooled tibiae from left and right bones in each cage is plotted in Fig. 3. In this experiment the largest dose of 13.5 units/100 g. of basal ration

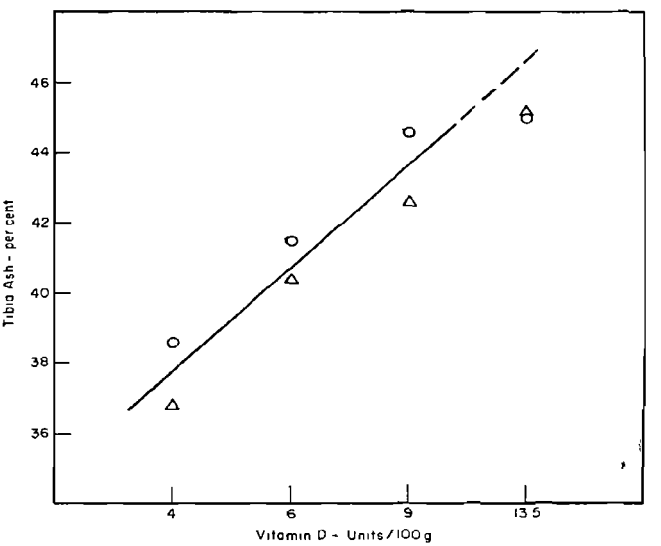

FIG. 3. Dosage-response curve for standard in chick assay of vitamin D in Table XVII, circles representing cages in battery I and triangles cages in battery II.

seemed to fall above the linear part of the dosage-response curve, so that at a 9-unit level the unknown feeding oils were at the upper end of the range instead of in the middle as intended. The unpredictability of the sensitivity of test chicks to vitamin D has long been a problem in this assay and frequently leads to discarding the results at the largest or smallest dose of the standard. However, both the negative controls and the cages at 13.5 units of standard were included with the usable standards and all the unknowns in the analysis of variance in Table XVIII.

TABLE XVIII. Analysis of Variance of Vitamin D Assays in Table XVII

Term	D.F.	Sum of squares	Mean square	F
Between batteries	1	4.10	4.10	5.47
Slope, 4 to 9 units	1	34.81	34.81	46.48
Curvature (Q^2)	1	.12	.12	.16
Remainder for treatments	8	193.77		
Error (batteries $\times$ treatments)	10	7.49	.749	1.00
Total	21	240.29		
Correction for mean	1	2581.22		

This has been calculated as an experiment in two randomized groups, essentially with the work form in Table X. The two batteries differed significantly in response, so that the segregation of differences between batteries has reduced the error and increased the precision of the assay. The slope was computed with $x_1 = -1$, 0, and 1 for doses 4, 6, and 9 units respectively as shown beneath Table XVII; it was highly significant. Curvature in this dosage range could be tested with Q^2, computed by Eq. 11; it was less than the error and clearly negligible.

The calculation of the assayed log potency of each oil by Eq. 23 is shown in the right side of Table XVII, the first entry being $M = .778 - .954 + (11.75 - 10.75)/16.75 = -.116$. The others were obtained similarly. The antilogarithm of each M, multiplied by 100, is its assayed potency as per cent of claim.

B. ASSAYS FROM TWO DOSAGE-REPONSE CURVES

The most general type of assay is that from two dosage-response curves, one for the standard and the other for the unknown, in which there is no limitation on the distribution and number of doses of the two preparations. A preliminary step is to discard any observations at the ends of the curve which on graphic analysis fall clearly outside the linear zone of response.

With two curves it is possible to test the validity of measuring the unknown in terms of the standard. If the preparations are qualitatively similar, the slopes of their curves should agree within the sampling error. If not, the response to one preparation may be influenced by dietary factors not present in the other or the preparations may differ qualitatively. The relative potency computed from diverging curves will differ with the level of response, and the unknown cannot be assigned a single potency. But if the curves are substantially parallel, the slope of both the standard and the unknown can be considered as separate estimates of the same parameter β, and the data can be pooled to obtain a more reliable estimate than that of either preparation.

The first stage is to compute two dosage-response curves, one for the standard and another for the unknown, as described in section II. The variability of the original observations around their respective curves will usually agree within the experimental error, and this condition is assumed here. Then the combined slope b_c of the two lines is determined from the numerators and the denominators of the individual slopes for the standard (s) and for the unknown (u) as

$$b_c = \frac{[xy]_s + [xy]_u}{[x^2]_s + [x^2]_u} = \frac{S[xy]}{S[x^2]} \tag{24}$$

When the individual curves have the same denominators, the combined slope is equal to the average of their slopes; in other cases it represents a weighted average. The total variation in y accounted for by the combined slope is

$$B_c^2 = \frac{S^2[xy]}{S[x^2]} \tag{25}$$

The significance of the difference between the component slopes depends upon the sum of squares between slopes, which with one degree of freedom is computed as

$$[y^2]_b = B_s^2 + B_u^2 - B_c^2 \tag{26}$$

The variance from Eq. 26 is then compared with the pooled variation about the two individual curves,

$$s^2 = \frac{[y^2]_s - B_s^2 + [y^2]_u - B_u^2}{n_s + n_u} \tag{27}$$

where subscripts s and u refer to standard and unknown respectively, and the two components of variation agree within the sampling error. The variance ratio $F = [y^2]_b/s^2$, with $n_1 = 1$ and $n_2 =$ degrees of freedom in s^2, tests the significance of a difference in slope. The above procedure can be arranged in the form of an analysis of variance, as is shown in Table XXI.

Three or more curves can be tested for parallelism by these same methods. The combined slope and the variation it accounts for are computed without change from the second form of Eq. 24 and from Eq. 25. The sum of squares measuring the divergence in slope can be written as

$$[y^2]_b = S(B_i^2) - B_c^2 \tag{26a}$$

where B_i^2 is the value computed by Eq. 9 for each curve. The degrees of freedom in the sum of squares among slopes is equal to the number of

individual slopes less one. For the test of significance the resulting mean square is divided by s^2, computed as

$$s^2 = \frac{S[y^2]_i - S(B_i{}^2)}{S(n_i)} \tag{27a}$$

where, as before, the subscript i designates the individual curves in the series.

In a typical assay individual observations for standard and unknown can be fitted by two parallel straight lines. The horizontal distance between them is measured in terms of the log dose and depends upon the relative potency of the two preparations. The log ratio of potency M may be computed by a slightly different form of Eq. 23 as

$$M = \bar{x}_s - \bar{x}_u - \frac{\bar{y}_s - \bar{y}_u}{b_c} = \bar{x}_s - \bar{x}_u + \frac{\bar{y}_u - \bar{y}_s}{b_c} \tag{23a}$$

where the $\bar{x}$'s and $\bar{y}$'s are the means of the individual curves computed by Eqs. 2 and 3. If the assumed doses of the unknown are the same as those of the reference standard, so that $\bar{x}_u$ is equal to $\bar{x}_s$, the equation simplifies to

$$M' = \frac{\bar{y}_u - \bar{y}_s}{b_c} \tag{23b}$$

The antilogarithm of M or of M' gives the potency of the unknown relative to the standard. To obtain (in logarithms) the units of vitamin per gram of the unknown, the logarithm of the assumed unitage of the unknown is added to M.

Example 7a. An assay from two dosage-response curves may be illustrated by data comparing crystalline vitamin D_3 with the reference cod liver oil in chicks. The per cent ash of the extracted tibia is shown in Table XIX. In this experiment the A.O.A.C. basal diet was modified

TABLE XIX. Assay of Vitamin D_3 (U) against Reference Cod Liver Oil (S)
Data from Lab. No. 1, Waddell and Kennedy (23)

Dose Units/100 g.	% Ash (— 30) S	U	Log dose
0	.97		
5.0	3.40	4.40	.699
7.5	7.28	5.95	.875
10.0	7.30	8.26	1.000
12.5	8.95	11.28	1.097
15.0	11.05	12.70	1.176
17.5	12.80	14.66	1.243
20.0	13.80	15.70	1.301
30.0	15.10	16.38	1.477
Total	79.68	89.33	8.868

and the chicks were depleted for 10 days before the test. In preparing the test diets the U.S.P. reference cod liver oil had an assigned potency of 115 units per gram and the crystalline vitamin D_3 an assumed potency of 45 million units per gram. The response at each dose was the percentage ash in a composite sample of tibiae from the 18 to 20 birds in each cage.

The calculation of the dosage-response curves for the standard and the unknown from the data in Table XIX by Eqs. 1 to 6 is shown in the first eight rows of Table XX; the entries in the last three rows were

TABLE XX. Computation of the Dosage-Response Curve for Standard and Unknown from the Data in Table XIX

Statistic	Standard	Unknown	Both
N	8	8	
$S(x)$	8.868	8.868	
$\bar{x}$	1.1085	1.1085	
$S(y)$	79.68	89.33	
$\bar{y}$	9.960	11.16625	
$[x^2]$	.429612	.429612	.859224
$[xy]$	6.70117	7.67025	14.37142
b	15.59819	17.85390	16.72605
B^2	104.5261	136.9439	240.3770
$[y^2]$	108.7306	143.7474	
$[y^2] - B^2$	4.2045	6.8035	11.0080

computed with Eqs. 8 and 9. The slope of the unknown was 14% steeper than that of the standard. To determine whether this difference fell within the experimental error, the numerators and denominators have been added in the last column to compute the combined slope and the variation attributable to it by Eqs. 24 and 25.

TABLE XXI. Analysis of Variance from the Data in Table XIX

Term	D.F.	Sum of squares	Mean square	F
Standard vs. unknown	1	5.8202	5.8202	6.34
Effect of combined slope	1	240.3770	240.3770	262.05
Divergence in slopes of S and U	1	1.0930	1.0930	1.19
Variation about both curves	12	11.0080	.9173	1.00
Variation about standard curve	6	4.2045	.7007	1.00
Variation about unknown curve	6	6.8035	1.1339	1.62

The tests of significance have been summarized in the analysis of variance of Table XXI. The variability about the two straight lines has been compared in the last two rows. Since either could be considered as a sampling error, the variance ratio F has been computed by dividing

the larger mean square with n_1 degrees of freedom by the smaller mean square with n_2 degrees of freedom. In consequence, the resulting ratio ($F = 1.62$) is always larger than 1 and the probabilities in tables of F are doubled, values at the 5% level of significance being read from the table for $P = .025$. In this case the two error variances agreed very well, so that they have been pooled to obtain $s^2 = .9173$ for the assay. The variance ratio for the divergence in slope was $F = 1.19$, which is also very much smaller than that required for significance. The highly significant F for slope completed the requirements for a valid assay.

In this experiment the same dosage levels were used for both standard and unknown. In consequence, the mean response to the standard and to the unknown could be compared by computing an additional row in the analysis of variance. The required sum of squares with 1 degree of freedom can be calculated from the difference between the totals $S(y)$ for the two preparations as $(89.33 - 79.68)^2/16 = 5.8202$, as shown in the first row of Table XXI. The F ratio indicated a significant difference ($P < .05$) between the assumed and assayed potencies of the crystalline vitamin D_3 or unknown. Since $\bar{x}_s = \bar{x}_u = 1.1085$, the log ratio of potency has been computed from Eq. 23b as $M' = (11.166 - 9.960)/16.726 = .07212$. The unknown was assumed to have a potency of 45 million units per gram, its logarithm being 7.6532. Adding M' to the logarithm of the assumed potency, we have 7.7253 as the logarithm of the assayed potency. Its antilogarithm shows an assayed potency of 53.13 million units per gram.

2. Factorial Determinations of Potency

One of the most important experimental techniques is the factorial design (19). A biological response depends upon the combined action of many variables, and the controlled experiment has long been relied upon for untangling their several effects. In studying a growth response, for example, the age and weight of the test animal, its environment, the dietary regime of its mother, the composition of its diet, and the manner of feeding are all factors which may modify the result. The traditional approach is to vary only one factor at a time and meanwhile to hold all others constant. In consequence, each factor is tested against a single background, and at the end of an experiment the investigator does not know to what extent his results depend upon the combination of factors he has used.

In contrast, a factorial experiment varies several factors simultaneously, so that each is tested over a wide range of known backgrounds. If the differences in response between two or more levels of a dietary ingredient recur at several levels of the other ingredients, under varying

environments, and with different methods of feeding, we have a much broader basis of inductive inference than if this information were lacking. More important, a change in one factor may have a marked effect upon the response to a second factor, and with the older doctrinaire approach this important result would be missed altogether. Despite the many advantages of factorial experiments in other aspects of nutritional research, we will apply the method here only to the problem of assigning doses in a determination of potency.

A. THE FACTORIAL DESIGN IN BIOLOGICAL ASSAYS

In assigning the doses of a biological assay, the two principal factors are preparation and dosage level. Usually only two preparations are compared, an unknown and a standard. Even when several unknowns are included in the same assay, each is compared independently with the same standard. Assays comparing several preparations in all possible combinations, such as the international collaborative assay for vitamin D (24), are exceptional and will not be considered here. Dosage level is subject to more variation. Although the simplest assays are those with only two dosage levels, many occur with three or more, and these fit readily into the standard factorial designs In every case successive doses are separated by the same dosage interval in logarithms, both with the standard and the unknown.

It is customary to label factorial designs by the number of factors and the number of levels of each. Thus a $2 \times 3 \times 2$ experiment might compare the growth response of rats to vitamin A alcohol and vitamin A ester, each at three dosage levels and with two different basal rations, requiring $2 \times 3 \times 2 = 12$ different treatment combinations. It is obvious that the third factor in such an experiment could be divided into many more, but in factorial assays we will confine our attention to the $2 \times k$ designs, where two preparations, a standard and an unknown, are each tested at k corresponding dosage levels.

The most useful designs are the 2×2 and 2×3 assays, requiring four and six treatment combinations respectively. Both provide an estimate of the slope of the dosage-response curve and test whether it is the same for both standard and unknown. Although factorial designs are concerned only with the selection of the doses in an assay, they are often combined with randomized groups, Latin squares, and other arrangements having a similar objective. When the experimental material can be arranged in relatively homogeneous groups of six, the 2×3 assay provides in addition a test of the linearity of the dosage-response curve which may be invaluable. Groups of six, however, may be more difficult to assemble than those of four, as in assays requiring litter mates of the

same sex. In these and other cases, the increased precision within groups is available with the 2×2 design. In all factorial experiments the same number of animals or potential responses must be assigned to each treatment combination either entirely at random or with a randomized design.

The results of factorial experiments are analyzed so as to isolate and measure the effect of each treatment factor individually. This is done most readily with factorial coefficients in a characteristic work form. Since this depends largely upon the number of doses, the analysis of 2×2, 2×3, and larger assays will be described separately.

B. TWO-DOSE FACTORIAL ASSAYS

The initial study of the dosage-response curve may have shown that the experimental error can be reduced by suitable restrictions in design. If we also know the range within which the response is a linear function of the log dose, the 2×2 factorial design is well adapted to homogeneous groups of four responses. The four treatment combinations may be designated as S_1, S_2, U_1, and U_2, representing the low and high doses respectively of the standard and of the unknown. It is essential that $S_2/S_1 = U_2/U_1$, so that the difference between the logarithms of the two dosage levels is constant. It is also essential that the number of responses at each dosage level (f) be the same. If in advance of treatment the material has been arranged in homogeneous groups, the entire record can be analyzed as an experiment in randomized groups or in Latin squares. The error term is then the interaction of treatments by restrictions in design, and this part of the calculation follows the same form as Table X or XIV. Otherwise the error term consists of the deviations about the dose means as in Table VI.

When the experiment is completed, the f response metameters are totaled for each of the four treatment combinations. Factorial analysis is concerned with isolating from the four total responses the three treatment comparisons corresponding to the two factors in the experiment and their interaction. This is accomplished most readily by factorial coefficients, which will also be designated by the letter x, and the work form in Table XXII. The factorial coefficients in each row total zero, and the products of the corresponding coefficients in any two rows also total zero. When both of these conditions are satisfied the comparisons are independent of one another and are said to be "orthogonal." When the number of orthogonal comparisons is equal to the degrees of freedom between doses or treatments, the sum of the variances computed with the coefficients in each row must account completely for the variation between treatments as measured by the sum of squares for doses or treatments in the original analysis of variance.

TABLE XXII. Work Form for Separating the Treatment Factors a and b and their Interaction ab in a Two-Dose Factorial Assay

Variance due to	Factorial coefficients (x) for dose				Divisor $fS(x^2)$	Sum of products $S(xT_d)$	Variance $S^2(xT_d)$ / $fS(x^2)$
	S_1	S_2	U_1	U_2			
(a) Difference between standard and unknown	−1	−1	+1	+1	$4f$	T_a	D^2
(b) Combined slope of dose-effect curve	−1	+1	−1	+1	$4f$	T_b	B^2
(ab) Divergence in slope	+1	−1	−1	+1	$4f$	T_{ab}	$[y^2]_b$
Total of f responses on each dose $= T_d$							

In Table XXII the first row (a) measures the difference in response between the standard and the unknown at corresponding dosage levels. The second row (b) provides an estimate of the combined slope of their dosage-response curves from the difference between the high and the low dosage levels. The third row (ab) tests whether the slopes of the individual curves differ significantly; its coefficients are the products of those in rows a and b. The sums of the responses (T_d) from the four doses, S_1, S_2, U_1 and U_2, are entered in the last row of the work form. The coefficients show how these totals are combined in each row to obtain the sums of products (T_a, T_b, and T_{ab}) in the next to the last column of the table. T_a is a function of the difference between two dosage-response curves in position (a) or of $a_u - a_s$, T_b leads directly to their common regression coefficient b_c, and T_{ab} measures the interaction of a and b in the factorial design.

The "divisor" for each row is equal to the sums of the squares of its coefficients multiplied by the number of responses at each dosage level or the frequency f. It is required in testing the significance of each sum of products either by the t test or the variance ratio F. For the t test each sum of products T_i, such as T_a, T_b, and T_{ab}, is divided by its standard error to obtain

$$t = \frac{T_i}{\sqrt{s^2 fS(x^2)}} \tag{28}$$

where s^2 is the mean square for error in the analysis of variance and $fS(x^2)$ is from the same row of the work form as T_i. The observed t is referred to a table of t (1) with the degrees of freedom in s^2 to determine whether a given T_i differs significantly from zero. Alternatively, the variance for each treatment factor may be computed by squaring its

sum of products and dividing by the corresponding divisor, $S^2(xT_d)/fS(x^2)$, as indicated in the last column of the work form. The ratio of each variance to s^2 is the variance ratio F, with which its significance can be tested as in an analysis of variance. The sum of the three treatment variances should equal exactly the sum of squares for treatments with three degrees of freedom in the original analysis of variance. This checks the correctness of the arithmetic.

The tests for the significance of T_t or of the treatment variances in the last column of Table XXII determine the validity of the assay. If the standard and unknown differ significantly in slope (T_{ab} significant) but not in their mean response (T_a not significant), a qualitative difference in the two preparations is indicated. If, however, they differ significantly in both slope (T_{ab}) and position (T_a), the dosage levels for one of the preparations may have been outside the central linear part of the dosage-response curve where it approaches an upper or lower limit of response. In the latter case it is advisable to revise one's estimate of the potency of the unknown and repeat the assay.

The relative potency in logarithms of assumed units of the unknown may be computed from a factorial assay with two doses as

$$M' = \frac{iT_a}{T_b} \tag{29}$$

where i is the interval between successive log doses and T_a and T_b are from Table XXII. The antilogarithm of M' measures the proportionate potency of the unknown in terms of the assumed potency. It is converted to original units by multiplying by the assumed unitage per gram.

Example 4c. The calculation of the factorial two-dose assay may be illustrated by the experiment on vitamin A in Table XI. Instead of considering this as a single dosage-response curve, let us assume that the first and third dosage levels represented two doses of an unknown and the second and fourth two doses of the standard, as indicated in the last column of the table. Within preparations the logarithm of the dosage interval $i = \log (2) = .3010$. In computing the "assay" the unknown is assumed to have the same potency as the standard, although its true potency is known to be the antilogarithm of $-.1505$ or 70.7%. We can determine whether the assayed potency is consistent with its true value.

The dose totals in Table XI have been substituted at the bottom of a work form such as that in Table XXII to compute T_a, T_b, and T_{ab} as shown in Table XXIII. Each of these was squared and divided by $4f = 48$ to obtain the variances in the last column. In comparison with the error term ($s^2 = 10.09$), the unknown differed significantly from the

TABLE XXIII. Factorial Analysis of Two-Dose "Assay" for Vitamin A from the Gains per Week of Female Rats in Table XI

Variance due to	Factorial coefficients (r) for S_1	S_2	U_1	U_2	Divisor $4f$	Sum of products $S(xT_d)$	Variance $S^2(xT_d)/48$
(a) Standard vs. unknown	−1	−1	1	1	48	$-98.6 = T_a$	$202.54 = D^2$
(b) Combined slope	−1	1	−1	1	48	$113.0 = T_b$	$266.02 = B^2$
(c) Lack of parallelism	1	−1	−1	1	48	$-5.8 = T_{ab}$	0.70
Dose total T_d	54.5	113.9	8.1	61.7			$10.09 = s^2$

standard and the high dose from the low dose but the dosage-response curves for standard and unknown were parallel. Hence we may compute the log ratio of potencies by Eq. 29 as $M' = (.3010)(-98.6)/113.0 = -.2626$ where $-.1505$ is expected. In original units, the assay indicated a potency of 54.6% although the true value was 70.7%. In Example 4d we will test whether this discrepancy is within the variation to be expected from the experimental error.

The three variances for the "assay" in Table XXIII totaled 469.26. This may be compared with the sum of squares for slope plus curvature of $439.02 + 30.24 = 469.26$ from the earlier analysis in Table XII of the variation among the same four dose totals. The two analyses started with different assumptions as to the interrelations among doses, one that they formed a single dosage-response curve and the other that they constituted a two-dose assay. Since each hypothesis accounted fully for the three degrees of freedom among the four totals, they gave necessarily the same total sum of squares.

C. THREE-DOSE FACTORIAL ASSAYS

The design of three-dose factorial assays is the same as that for two doses except that both the standard and the unknown are administered at three dosage levels spaced equally on a logarithmic scale. In addition to the three comparisons of the two-dose assay, the three-dose design tests directly the hypothesis of a linear relation between log dose and response. The first three rows of Table XXIV have the same meaning as the corresponding rows in Table XXII. The two additional terms are computed by the factorial coefficients in the last two rows. Each sum of products (T_i) in the last column of Table XXIV may be tested for significance by Eq. 28.

Most graded response assays are based for convenience upon a linear segment of a dosage-response curve which over a wider dosage range would be sigmoid in form. It is especially desirable, therefore, to deter-

TABLE XXIV. Work Form for the Factorial Analysis of a 2×3 Assay

Variance due to	Factorial coefficients (x) for dose						Divisor $fS(x^2)$	Sum of products $S(xT_d)$
	S_1	S_2	S_3	U_1	U_2	U_3		
(a) Differences between samples	−1	−1	−1	+1	+1	+1	$6f$	T_a
(b) Slope of dosage-response curve	−1	0	+1	−1	0	+1	$4f$	T_b
(ab) Departure from parallelism	+1	0	−1	−1	0	+1	$4f$	T_{ab}
(c) Curvature of combined curve	+1	−2	+1	+1	−2	+1	$12f$	T_c
(ac) Opposed curvature of separate curves	−1	+2	−1	+1	−2	+1	$12f$	T_{ac}

mine in each case that the selected doses in fact can be treated as if they defined a straight line. The test for curvature is based upon the quadratic term of a parabola as computed with the orthogonal polynomials (x_2) described on p. 467. The fourth row of the table measures curvature in the same direction in the curves for both the standard and the unknown and the fifth row curvature in opposite directions from the interaction of the factors in rows a and c. A significant combined term suggests that the test doses fall outside the linear zone or too near the "ceiling" or the "floor" for both preparations, especially if the variance between preparations is not significant. A significant last row may indicate too large a dosage interval or, if the difference between samples is also significant, too great a discrepancy between the assumed and the true potency of the unknown. In either case it may be possible to salvage the assay by omitting the responses at dosage levels which fall outside the central linear portion of the complete curve.

For a satisfactory factorial three-dose assay none of the T_i's in the last three rows of Table XXIV should be significant, and their mean square should not approach significance. If the assay meets these requirements for validity, the relative potency of the unknown may be computed in logarithms of assumed units as

$$M' = \frac{4iT_a}{3T_b} \tag{30}$$

which has the same interpretation as in the two-dose assay.

Example 8a. The three-dose assay may be illustrated by data reported by Coward and Kassner (25) on a vitamin D line test. The degree of healing in the split tibia of each rat was scored on an arbitrary scale, which has here been multiplied by 4 to eliminate fractions. Table XXV gives the response for one rat in each of 12 litters at three different dos-

age levels of two samples of irradiated ergosterol, one called the "standard" and the other the "unknown." The sums of squares between

TABLE XXV. Assay of Vitamin D Oil in Terms of the Degree of Healing (Line Test) in Which the Original Grades of Healing Have Been Multiplied by 4

Data of Coward and Kassner (25)

Litter	Standard, mg.			Unknown, mg.			
No.	2.5	5	10	2.5	5	10	T_g
1	2	8	8	3	9	7	37
2	6	4	9	3	5	8	35
3	4	6	12	4	6	9	41
4	9	11	10	6	14	13	63
5	10	15	17	8	8	10	68
6	7	7	5	6	9	9	43
7	4	10	13	5	11	13	56
8	11	4	9	3	6	15	48
9	2	9	14	5	8	6	44
10	4	7	13	4	10	10	48
11	12	10	9	15	18	15	79
12	4	8	11	7	8	12	50
T_d	75	99	130	69	112	127	612

litters, between doses and the interaction of doses by litters which are shown in Table XXVI have been computed from the marginal totals and the individual entries of Table XXV. The calculation followed the work form for randomized groups in Table X except that the differences be-

TABLE XXVI. Analysis of Variance for the Vitamin D Assay in Table XXV

Variation	D.F.	Sum of squares	Mean square	F
Between litters	11	321.00	29.18	4.04
Between doses	5	278.00		
Litters $\times$ doses	55	397.00	$7.22 = s^2$	1.00
Total	71	996.00		
Correction for mean	1	5202.00		

tween dosage levels have not been subdivided. It is evident that the segregation of litter differences increased the sensitivity of the assay materially. If the variation between litters were included in the error, half again as many animals would be needed to obtain the same precision.

The variation in response between different dosage levels was divided into the five components of biological interest in the factorial analysis of Table XXVII. The variance attributable to each term in the last column

TABLE XXVII. Factorial Analysis for the Three-Dose Assay in Table XXV

Effect	Factorial coefficients (x) for dose						Divisor	Sum of products	Variance
	S_1	S_2	S_3	U_1	U_2	U_3	$fS(x^2)$	$S(xT_d)$	$\frac{S^2(xT_d)}{fS(x^2)}$
S vs. U	—1	—1	—1	1	1	1	72	4	$.22 = D^2$
Slope	—1	0	1	—1	0	1	48	113	$266.02 = B^2$
Parallel	1	0	—1	—1	0	1	48	3	.19
Curvature	1	—2	1	1	—2	1	144	—21	3.06
Opposed curvature	—1	2	—1	1	—2	1	144	—35	8.51
T_d	75	99	130	69	112	127		Error	$s^2 = 7.22$

may be compared with $s^2 = 7.22$ from Table XXVI. Since the difference between standard and unknown was much less than the experimental error, the results of the assay were consistent with the original assumption that the two samples had the same potency. The variance for the combined slope was very highly significant, indicating an adequate assay technique, while the deviations from parallelism and from linearity were well within the experimental error. Since the data met the requirements for a valid assay, the relative potency in logarithms could be computed by Eq. 30 as

$$M' = \frac{4 \times .3010 \times 4}{3 \times 113} = 0.0142$$

The potency of the unknown, therefore, was assayed as 1.033 times that of the standard.

Example 9a. A biological assay of the vitamin C activity of fresh orange juice may be used as a second example. The response of 10 guinea pigs on each of three doses of ascorbic acid (S) and of fresh orange juice (U) during a six-week test period was measured by Crampton (26) from the average length of the odontoblasts in each animal (p. 257), with the results given in Table XXVIII. Five of the animals at each treatment combination were males and five were females but except for this sex restriction the guinea pigs were assigned to doses entirely at random. The assay could be considered, therefore, as forming a $2 \times 3 \times 2$ factorial experiment and has been so analyzed. The complete analysis (not given here) showed that the effect of sex could be neglected, so that in the interests of simplicity the data will be treated here as a 2×3 assay with 10 responses at each dose.

TABLE XXVIII. Biological Assay of Vitamin C Activity of Orange Juice from the Length of Odontoblasts in Incisors of Guinea Pigs

Data of E. W. Crampton (26).

Sex	Odontoblast readings ($\mu - 20$) for treatments					
	Ascorbic acid (S)			Fresh orange juice (U)		
	S_1	S_2	S_3	U_1	U_2	U_3
Males	4.2	16.1	23.6	15.2	19.7	25.5
	11.5	16.1	18.5	21.5	23.3	26.4
	7.3	15.2	33.9	17.6	23.6	22.4
	5.8	17.3	25.5	9.7	26.4	24.5
	6.4	22.1	26.4	14.5	20.0	24.8
Females	10.0	17.3	32.1	10.0	25.2	30.9
	11.2	13.6	26.7	8.2	25.8	26.4
	11.2	14.5	21.5	9.4	21.2	27.3
	5.2	18.8	23.3	16.1	14.5	29.4
	7.0	15.5	29.1	9.7	27.3	23.0
Total	79.8	166.5	260.6	131.9	227.0	260.6
Mean	7.98	16.65	26.06	13.19	22.70	26.06
$[y^2]$	67.90	53.13	199.64	176 53	137.66	63.44

The fresh orange juice or unknown was analyzed chemically for its ascorbic acid content and on this basis fed without alteration at the same three levels of ascorbic acid as the standard. The ascorbic acid standard was administered at 0.5, 1.0 and 2.0 mg. per day in a mixture of corn starch and corn meal. Hence the objective of the assay was to determine whether under these conditions the antiscorbutic activity of the orange juice was greater than would be predicted from its content of ascorbic acid as determined chemically in the present experiment.

TABLE XXIX. Factorial Analysis of Three-Dose Assay in Table XXVIII

Variance assigned to	Factorial coefficients (x) for dose						Divisor	Sum of products	Variance T_i^2
	S_1	S_2	S_3	U_1	U_2	U_3	$fS(x^2)$	$S(xT_d) = T_i$	$fS(x^2)$
(a) Standard vs. unknown	−1	−1	−1	1	1	1	60	112.6 = T_b	211 31[a]
(b) Combined slope	−1	0	1	−1	0	1	40	309.5 = T_b	2394.76[a]
(ab) Lack of parallelism	1	0	−1	−1	0	1	40	−52 1 = T_{ab}	67 86[b]
(c) Combined curvature	1	−2	1	1	−2	1	120	−54.1 = T_c	24.39
(ac) Opposed curvature	−1	2	−1	1	−2	1	120	−68 9 = T_{ac}	39 56
Dose total T_d	79 8	166.5	260 6	131 9	227 0	260 6		Error s^2 =	12 931

[a] Significant at $P < .001$.
[b] Significant at $P < .05$.

The work form for the three-dose factorial assays in Table XXIV has been used in Table XXIX to isolate the T_i's corresponding to the five degrees of freedom for treatments. The sum of squares has been computed separately from the responses for each treatment (Eq. 8) as listed in the last row of Table XXVIII. Since the $[y^2]$'s at the different dosage levels did not differ more than would be expected by chance, the error variance has been computed from $S[y^2] = 698.30$ with $6 \times 9 = 54$ degrees of freedom to obtain $s^2 = 12.931$. To test the significance of each sum of products, the variance in the last column of Table XXIX was divided by s^2 to obtain the variance ratio F. The dosage-response curves for S and U diverged significantly and both terms for curvature exceeded their errors. It is evident that the potency of the unknown could not be computed as in a valid three-dose assay.

The source of the discrepancy was not difficult to see when the data were plotted in Fig. 4. The response to the highest dose of the unknown

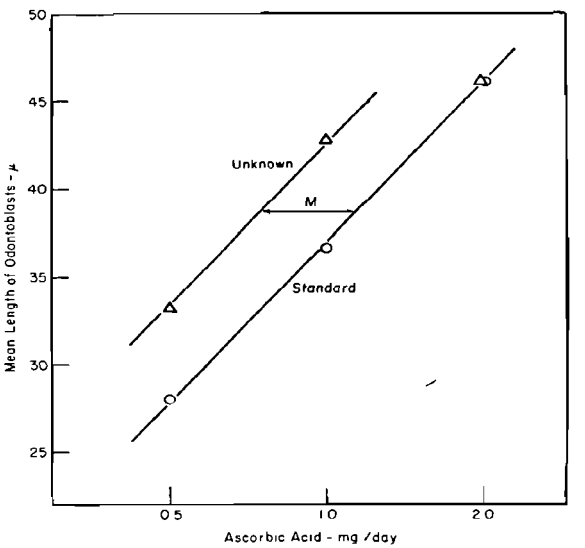

FIG. 4. Odontoblast assay of vitamin C in guinea pigs from data in Table XXVIII.

was up against a "ceiling." By omitting U_3, the remaining data could be fitted by parallel straight lines and M computed from the horizontal distance between them. The slope for the standard was computed from all three dosage levels with $x_1 = -1$, 0, and 1 and that for the unknown

from the two lower dosage levels with $x_1 = -1$ and 1. Since these were to be combined in determining b_c, they were calculated by Eqs. 4c and 5c as

	Standard		Unknown		$S + U$
$[x^2]$	$.301^2 \times 10 \times 2$	= 1.81202	$(.301/2)^2 \times 10 \times 2$	= .45300	2.26502
$[xy]$	$.301(260.6 - 79.8)$	= 54.4208	$.301(227.0 - 131.9)/2$	= 14.3126	68.7334
b	30.0332		31.5951		30.3456
B^2	1634.432		452.209		2085.757
$\bar{x}$	0		—.1505		
$\bar{y}$	16.897		17.945		

The two lines were satisfactorily parallel, since $[y^2]_b = 1634.432 + 452.209 - 2085.757 = .884$, which was less than the revised error term $(-U_3)$, $s^2 = 14.108$ with 45 degrees of freedom. The curve for the standard was also satisfactorily linear, since $Q^2 = (79.8 - 2 \times 166.5 + 260.6)^2/(6 \times 10) = .913$, also less than s^2. Substituting in Eq. 23a, $M = .1505 + .0345 = .1850$ for the log potency of the orange juice. Hence the antiscorbutic activity of the orange juice was (antilog .1850) = 1.53 that of its ascorbic acid content. By biological assay its potency was more than half again as large as that predicted from the chemical tests used in this experiment.

D. FACTORIAL ASSAYS WITH MORE THAN THREE DOSES

Factorial assays with more than three doses sometimes occur. With the larger number of doses, nonlinearity in the dosage-response curves can usually be recognized from the trend of the plotted points and only those responses retained which fall in the linear range. The first three factorial comparisons are isolated individually, but the terms for curvature are combined into a single value measuring the scatter of the dose means about the two straight lines as in Table VI. If this scatter exceeds the variation between replicates significantly, it is the appropriate error of the assay.

The factorial coefficients are equivalent to those in the first three rows of Tables XXII and XXIV. The total response on the standard is subtracted from that on the unknown to obtain T_a. The coefficients for computing the slope are the x_1's on p. 458 and depend upon whether the number (k) of dosage levels is odd or even. The sum of products, $S(x_1T_d)$, is computed separately for the standard and for the unknown. These two sums are added to obtain $T_b = S(x_1T_d)_s + S(x_1T_d)_u$ and one is subtracted from the other to obtain $T_{ab} = S(x_1T_d)_u - S(x_1T_d)_s$. The variances D^2, B^2, and $[y^2]_b$ are computed from T_a, T_b, and T_{ab} as shown in the preceding sections.

The log ratio of potency M is computed by Eq. 23b, in which the

combined slope is determined from T_b by Eq. 7 with $T_b = S(x_1T_d)$ and

$$\bar{y}_u - \bar{y}_s = \frac{2T_a}{N} \tag{31}$$

where N is the total number of observations.

TABLE XXX. Microbiological Assay for Riboflavin by the Method of Light and Clarke. Response is the Titer per Tube in Milliliters of 0.1 N NaOH

Data of P. R. Burkholder

Dose	Standard			Unknown			Coded dose
μg./tube	Tube 1	Tube 2	T_d	Tube 1	Tube 2	T_d	x_1
.283	4.85	5.00	9.85	4.10	4.50	8.60	—7
.400	6.20	6.40	12.60	5.50	5.60	11.10	—5
.566	8.25	7.70	15.95	7.45	7.00	14.45	—3
.800	8.50	8.50	17.00	7.90	8.80	16.70	—1
1.130	10.30	10.40	20.70	9.90	9.15	19.05	1
1.600	11.40	11.60	23.00	10.50	10.25	20.75	3
2.260	13.70	13.60	27.30	12.50	12.55	25.05	5
3.200	13.75	13.50	27.25	13.80	13.75	27.55	7
Total			153.65			143.25	
$S(x_1T_d)$			220.15			223.65	

$T_a = 143.25 - 153.65 = -10.40$ $\quad T_b = 220.15 + 223.65 = 443.80$

$T_{ab} = 223.65 - 220.15 = 3.50$

$fS(x_a^2) = 32$ $\quad fS(x_b^2) = fS(x_{ab}^2) = 672$

Example 10. The data from an eight-dose factorial microbiological assay of riboflavin are given in Table XXX. After excluding the doses which fell outside the linear portion of the curve, the latter still covered an ample dosage range of more than tenfold. The "unknown" was an 84.1% concentration of the standard. The duplicate tubes at each dose were prepared and handled together throughout the assay. Hence they might be expected to agree more closely with one another than their means with the two dosage-response curves, as could be tested by an analysis of variance.

TABLE XXXI. Analysis of Variance of Riboflavin Assay in Table XXX

	D.F.	Sum of squares	Mean square	F	
Standard vs. unknown	1	3.3800	3.3800	11.54 [a]	
Effect of combined slope	1	293.0929	293.0929	1001.00 [a]	
Divergence in slopes of S and U	1	.0182	.0182	.06	
Scatter of T_d about both curves	12	3.5136	.2928	1.00	4.07 [a]
Variation between duplicates	16	1.1500	.0719		1.00
Total	31	301.1547			
Correction for mean	1	2754.6753			

[a] $P < .01$.

The total variation among the 32 titers has been divided into five components by the analysis of variance in Table XXXI. The first three entries were determined from the values at the bottom of Table XXX for T_a, T_b and T_{ab} and their divisors. In computing the scatter of the dose totals or means about the two dosage-response lines, the sum of squares among the T_d's was calculated from the 16 totals for T_d as $(9.85^2 + 12.60^2 + \cdots + 27.55^2)/2 - 296.90^2/32 = 300.0047$, and the first three terms were subtracted to obtain 3.5136 with 12 degrees of freedom. The variation between duplicates was obtained as a difference from the sum of squares for all individual tubes. The significantly larger mean square for scatter than for the variation between duplicates occurs not infrequently in microbiological assays where duplicate tubes are handled together. This reduces drastically their potential value for increasing the precision of such assays. The experimental error was necessarily $s^2 = 0.2928$ with 12 degrees of freedom rather than 0719, the variance between duplicate tubes.

The assayed potency of the "unknown" depended upon two terms. The difference in the mean responses $\bar{y}_u - \bar{y}_s$ was computed from T_a by Eq. 31 as $2(-10.40)/32 = -.6500$. The combined slope was determined from T_b by Eq. 7 as $b = 443.80/(.07525 \times 672) = 8.7763$, since $I' = .1505/2 = .07525$. Substituting in Eq. 23b, $M' = -.6500/8.7763 = -.07406$, the true potency being log $(.841) = -.07520$.

3. The Precision of the Assayed Potency

Assays may vary enormously in precision, so that each determination of potency should be accompanied by an estimate of its reliability. This may be expressed in terms of its standard error or of its confidence or fiducial limits. These should indicate the agreement to be expected between independent similar assays of the same preparation, preferably in different laboratories. The variation among similar assays could be measured directly from the potencies reported by 6 or 8 collaborators of equal skill, but such estimates are rarely available. The error computed from the internal evidence of a single experiment should approximate one based upon independent replications. This can only be determined empirically. There may be sources of variation between assays which affect the estimate of potency but are not represented in the internal evidence of a single experiment. Occasionally these additional factors have been several times as large as those within a self-contained assay, but in many cases they have proved relatively unimportant. Judging from the available evidence, the intrinsic factors in the vitamin

assays are by far the most important. Usually they provide the only available measure of precision, which is expressed in terms of the standard or sampling error or of the more exact confidence or fiducial limits.

A. THE STANDARD ERROR OF POTENCY

The vitamin content of an unknown sample is computed in terms of M, its log potency relative to the standard. The standard error of this estimate s_M is computed similarly in logarithmic units. Composite values of the slope b and of the standard deviation s about the slope are determined from the dosage-response lines for the standard and for the unknown when they have been shown to agree within the sampling error. From s and b the standard deviation of the assay is determined in units of log dose by Eq. 16 as $\lambda = s/b$. The standard error of M in its most general form may then be written as

$$s_M = \lambda \sqrt{\frac{1}{N_u} + \frac{1}{N_s} + \frac{(\bar{y}_u - \bar{y}_s)^2}{B_c^2 - s^2t^2}} \tag{32}$$

where N_u and N_s are the total number of observations on the unknown and on the standard respectively and B_c^2 is computed by Eq. 25. The latter value is here diminished by s^2t^2, where t is taken from a table of t at $P = .05$ or .01 for the degrees of freedom in s^2. This refinement emphasizes the critical importance of designing the assay so that the slope will be unquestionably significant. If $B_c^2 - s^2t^2$ is negative, the assay is so lacking in precision that it has no determinate standard error. The desirability of this qualification in computing s_M is emphasized by Irwin's finding that in a series of 53 collaborative assays for vitamin A, 30% of them had indeterminate confidence limits at $P = .05$ (27). By defining s_M as in Eq. 32, this limitation would apply equally to the standard error. In about two out of three experiments the range represented by $M \pm s_M$ would be expected to include the true log potency of the unknown.

The standard error of a factorial assay may be calculated as

$$s_M = \lambda \sqrt{\frac{4}{N}\left\{1 + \frac{D^2}{B^2 - s^2t^2}\right\}} \tag{32a}$$

where D^2 and B^2 are the variances computed from T_a and T_b respectively, and N is the total number of observations. For a two-dose assay the combined slope b, required in computing λ, is readily determined as

$$b = \frac{T_b}{2if} \tag{7b}$$

and for a three-dose assay as

$$b = \frac{T_b}{4if} \tag{7c}$$

where f is the number of responses for each treatment.

The antilogarithms of $M + s_M$ and $M - s_M$ give upper and lower limits in original units. Some investigators prefer to report the antilogarithm of s_M instead, writing it after the potency with a $\times\!\!/$ sign. Alternatively, this may be expressed as the percentage standard error, which is equal to 100 (antilog s_M) $-$ 100. If the antilogarithm of s_M were 1.068, for example, the assayed potency would have a standard error of 6.8%. Because the $\pm$ is used so widely, however, the following less exact approximation for the standard error has had considerable vogue:

$$\text{S.E. of relative potency} = 2.303 s_M \text{ (antilog } M) \tag{33}$$

This approximation is reasonably satisfactory when s_M is small relative to M, but should not be used for standard errors larger than 10%.

For smaller odds than 1 in 3 of *not* bracketing the true value of M, such as 1 in 20 or $P = .05$, approximate limits may be computed from s_M. The standard error is multiplied by t and the product ts_M is added to M and subtracted from M. The required t is read from a table of t at the degrees of freedom (n) in the standard deviation s. The resulting limits are in logarithms and are then converted to original units. Approximate limits computed in this way have two limitations. They make insufficient allowance for the sampling error of the slope and they are centered around the most probable value M. Confidence or fiducial limits are but little more difficult to compute and are preferred in all critical cases.

Example 6b. The standard error of M with two responses at one dosage level of the unknown may be computed with Eq. 32 for the vitamin D chick assay in Tables XVII and XVIII. Here $s = \sqrt{.749} = .865$ with 10 degrees of freedom and $\lambda = .865/16.75 = .0516$. Since t at $P = .05$ and $n = 10$ is given by the t table (10) as 2.228, $B^2 - s^2t^2 = 34.81 - .749 \times 4.964 = 31.09$. Hence

$$s_M = .0516 \sqrt{\frac{1}{2} + \frac{1}{6} + \frac{(\bar{y}_u - \bar{y}_s)^2}{31.09}}$$

The unknowns A and E showed the smallest differences in $\bar{y}_u - \bar{y}_s$, giving $s_M = .0516 \sqrt{\frac{2}{3} + \frac{1}{31.09}} = .0431$ and unknown B with the largest differ-

ence had a standard error of

$$s_M = .0516 \sqrt{\frac{2}{3} + \frac{3.80^2}{31.09}} = .0549$$

The antilogs of the extreme values of s_M were 1.104 and 1.135 respectively, so that these control assays had standard errors ranging from 10.4 to 13.5%.

Example 4d. For computing the standard error of the two-dose factorial assay in Table XXIII for vitamin A, $b = 113.0/(2 \times .301 \times 12) = 15.64$ (Eq. 7b), $s = \sqrt{10.09} = 3.176$ with 32 degrees of freedom, $\lambda = 3.176/15.64 = .2031$ (Eq. 16) and $B^2 - s^2t^2 = 266.02 - 10.09 \times 2.038^2 = 224.11$ (with t for $P = .05$). Substituting in Eq. 32a,

$$s_M = .2031 \sqrt{\frac{4}{48}\left\{1 + \frac{202.54}{224.11}\right\}} = .2031 \times .3982 = .0809$$

giving a standard error of 20.5% of the assayed potency. The standard error of the percentage relative potency by Eq. 33 is here equal to

$$\text{S.E.} = 2.303 \times .0809 \times 54.6 = 10.2\%$$

so that at odds of 1 in 20 the true potency should fall within approximate limits of $54.6 \pm 2.038 \times 10.2 = 54.6 \pm 20.8 = 33.8$ and 75.4. The true potency in this case is known to be 70.7%.

Example 8b. The standard error of M for a three-dose factorial assay of vitamin D by the line test may be computed with Eq. 32a. From Tables XXVI and XXVII, $B^2 - s^2t^2 = 266.02 - 7.22 \times 7.12 = 214.61$ at $P = .01$, $b = 113/(4 \times .301 \times 12) = 7.821$ and $\lambda = 2.687/7.821 = .3436$. Substituting in Eq. 32a,

$$s_M = .3436 \sqrt{\frac{4}{72}\left\{1 + \frac{.22}{214.61}\right\}} = .0810$$

from which the percentage standard error was 100 antilog (.0810) $-$ 100 $=$20.5%.

B. CONFIDENCE OR FIDUCIAL LIMITS

Although they differ in basic theory, both confidence and fiducial limits for the log ratio of potencies are numerically the same. The importance of a significant slope has been emphasized in the preceding section by modifying the usual equation for s_M. As defined there, the effect of slope in the denominator has been reduced from B^2 to $B^2 - s^2t^2$ in accord with the equations for exact confidence limits. As a result of this adjustment the usual formulae for the confidence limits of biological assays (28) can be expressed in terms of s_M.

In the general case, the exact confidence limits X_L may be computed as

$$X_L = \bar{x}_s - \bar{x}_u - \frac{C^2(\bar{y}_s - \bar{y}_u)}{b_c} \pm tCs_M \tag{34}$$

where

$$C^2 = \frac{B^2}{B^2 - s^2t^2} \tag{35}$$

and s_M is computed by Eq. 32. C is always larger than 1, so that the exact confidence limits are not only wider than the approximate values computed with s_M alone, but the midpoint of the confidence interval does not coincide with the most probable value as given by M. Since confidence limits are equally spaced above and below the mean in terms of y, they are necessarily spaced unequally about M when projected to the x axis.

For factorial asays, the confidence limits are

$$X_L' = C^2M' \pm tCs_M \tag{34a}$$

where s_M is computed by Eq. 32a.

Example 7b. The computation of confidence limits may be illustrated with the data from the chick assay of vitamin D_3, where the unknown had an assumed unitage in logarithms of 7.6532. From the data in Tables XX and XXI

$$\lambda = \frac{\sqrt{.9173}}{16.726} = .05726$$

$$B^2 - s^2t^2 = 236.02$$

$$s_M = .05726\sqrt{2/8 + (9.960 - 11.166)^2/236.02} = .0290$$

Hence the assayed log potency was equal to $7.7253 \pm .0290$, and the potency was assayed with a standard error of 6.9%. The confidence limits for odds of 19 in 20 were computed with $t = 2.179$ at $P = .05$ and $n = 12$, and $C^2 = 240.38/236.02 = 1.0185$ by Eq. 35, leading by Eq. 34 to $X_L = 7.6532 + 1.0185 \times 1.2062/16.726 \pm 2.179 \times 1.0092 \times .0290 = 7.7266 \pm .0638 = 7.6628$ and 7.7904. The true potency presumably fell within limits of 46.0 and 61.7 million units per gram. Because of the relatively high precision of the slope the exact limits in this example agreed very closely with the approximate ones computed without C^2.

Example 4e. The confidence limits for the two-dose factorial assay of vitamin A in Table XXIII are an extension of the standard error of the log potency in Example 4d. By Eq. 35, $C^2 = 266.02/224.11 = 1.1870$ and $C = 1.0895$, leading by Eq. 34a to $X_L' = 1.1870\ (-.2626) \pm 2.038 \times 1.0895 \times .0809 = -.3117 \pm .1796 = -.4913$ and $-.1321$. Hence

the exact limits gave boundary values of 32.3 and 73.8% which differed but little from the approximate limits of 33.8 and 75.4%, where 70.7% was the true potency.

Example 9b. The calculation of confidence limits for the vitamin C assay in Table XXVIII may be summarized as follows:

$$B_c^2 - s^2t^2 = 2085.76 - 14.108 \times 2.016^2 = 2028.42$$

$$C^2 = 1.0283, \quad C = 1.0141 \qquad \text{(Eq. 35)}$$

$$\lambda = 3.756/30.346 = .1238 \qquad \text{(Eq. 16)}$$

$$s_M = .1238\sqrt{\frac{1}{20} + \frac{1}{30} + \frac{(17.945 - 16.897)^2}{2028.42}} = .03585 \qquad \text{(Eq. 32)}$$

$$X_L = .1505 + \frac{1.0283 \times 1.048}{30.346} \pm 2.016 \times 1.0141 \times .03585$$

$$= .1860 \pm .0733 = .1127 \text{ and } .2593 \text{ at } P = .05 \qquad \text{(Eq. 34)}$$

In percentage terms the antiscorbutic potency of orange juice was found to be 30 to 82% greater than would be expected from its ascorbic acid content as determined chemically.

4. Extensions of the Two-Dose Factorial Assay

Except that it lacks a direct test for curvature, the 2×2 factorial assay is usually the most satisfactory. When linked with randomized groups, its computation can be further simplified, two unknowns can be assayed against the same standard and with confounding it can be used for groups of two or three.

A. Abbreviated computation of assay results

The two-dose assay can be computed in a shortened form without a calculator when the experiment is arranged in randomized groups (29). The design of the assay and the computation of relative potency are unchanged, but the error variance is determined without an analysis of variance. The factorial coefficients for the two-dose assay are applied not only to the totals but also to the responses in each group in two steps. Four initial differences are determined from the responses of each group within dosage levels and within preparations, i.e., $D_1 = U_2 - S_2$, $D_2 = U_1 - S_1$, $D_3 = U_2 - U_1$, and $D_4 = S_2 - S_1$. By adding or substracting these differences in pairs, we obtain $y_a = D_1 + D_2$, $y_b = D_3 + D_4$, and $y_{ab} = D_1 - D_2$. Over all groups, these total to T_a, T_b, and T_{ab} respectively as previously defined. The log relative potency M is calculated from these totals by Eq. 29.

The principal advantage of the method is in the estimation of the error. Since the y_a's, y_b's, and y_{ab}'s are independent of the differences between group means, the error variance within groups (s^2) can be computed from the variation between the y's of each class. The efficient estimate is that based upon the squares of all the y's, pooling the variation within the three categories for a, b and ab. The error variance or square of the standard deviation is given by the equation

$$s^2 = \frac{S(y^2) - (T_a^2 + T_b^2 + T_{ab}^2)/f}{12(f-1)} \qquad (36)$$

where f is the number of groups. The variance obtained by this equation is identical with the mean square in the error row of an analysis of variance. It has $3(f-1)$ degrees of freedom.

No numbers are squared in an alternative method which for many purposes provides nearly as good an estimate of s as the efficient statistic given above. It is based upon the mean range $\overline{R}$ in the y's. A range (R) is the difference between the smallest and the largest number in a series. The smallest y_a may be subtracted from the largest y_a to obtain one range and the process repeated to obtain the equivalent ranges for y_b and y_{ab}. The average of these three ranges is multiplied by a factor depending upon the number of y's in each series to obtain an estimate of the standard deviation s. Table XXXII lists the factors applying to the y's in the present design for series of 2 to 11 items.

TABLE XXXII. Coefficients for Estimating the Standard Deviation of a Normal Distribution as the Product of the Coefficient and the Mean Range. $\overline{R}_y$ is the Mean Range in y_a, y_b and y_{ab} as Defined on p. 509 and Illustrated in Table XXXIII. $\overline{R}_d$ and $\overline{R}_p$ are the Mean Ranges in d_1 and d_2 and in p_1 and p_2 as Defined on p. 514 and Illustrated in Table XXXIV

Based upon Table XXII in Pearson's Tables (32)

No. of entries in set or subset	Coefficients for determining estimated standard deviation from	
	$\overline{R}_y$	$\overline{R}_d$ or $\overline{R}_p$
2	.4431	.6267
3	.2954	.4178
4	.2429	.3435
5	.2150	.3040
6	.1973	.2790
7	.1849	.2615
8	.1756	.2484
9	.1683	.2381
10	.1625	.2298
11	.1576	.2229

The efficiency of the method falls off rapidly as the series is lengthened beyond 10 or 11 entries. In consequence, it is advantageous to subdivide the full series of randomized groups into equal sets of 3 to 9 in the order in which they occur in the record and determine the range of each y in all sets. The average range over all sets is multiplied as before by the appropriate factor in Table XXXII to obtain an estimate of the standard deviation. Recently tables have been published for tests of significance with the standard deviation based upon the average range (30, 31). These indicate that at $P = .05$ and to a lesser extent at $P = .01$ the usual t table can be used here instead, assigning the estimate $3(f - 1)$ degrees of freedom, as if computed by Eq. 36.

With the standard deviation estimated from the mean range, one can apply the tests of significance which validate the assay and compute the error of the assayed potency. The variance ratio or F has been used in the preceding section to test the significance of the variances computed from T_a, T_b, and T_{ab}. Since the standard deviation is here determined directly rather than as the square root of a variance, it is more convenient in the balanced two-dose assay to compute the ratio of T_a, T_b, and T_{ab} (or, more generally, T_t) to its standard error as

$$t = \frac{T_t}{2s\sqrt{f}} \tag{28a}$$

with $3(f - 1)$ degrees of freedom. Assays with a significant T_b and a nonsignificant T_{ab} are considered suitable for computing the log ratio of potencies M by Eq. 29. Its standard error s_M may be determined by Eq. 32a or from the T_t's directly as

$$s_M = \lambda \sqrt{\frac{1}{f}\left(1 + \frac{T_a^2}{T_b^2 - 4fs^2t^2}\right)} \tag{32b}$$

Example 11. The abbreviated computation may be illustrated with the data in Table XXXIII of a five-week vitamin A assay reported by Irwin (27). The initial differences, D_1 to D_4, were computed for each litter as shown in the middle part of the table and then combined into the y's in the last three columns. To check the calculation the procedure was repeated with the dose totals T_d to obtain T_a, T_b, and T_{ab} respectively, which should agree exactly with the column totals for the corresponding y's.

Two methods for computing the standard deviation s are shown below the table. One was to divide the litters or groups into two groups of five and determine the range for each group. The mean range $\bar{R}$ was multiplied by the factor in Table XXXII for a series of five to estimate the

standard deviation as $s = 28.667 \times .2150 = 6.163$. This may be compared with the efficient estimate from Eq. 36 of $s = \sqrt{41.192} = 6.418$ with $3(10 - 1) = 27$ degrees of freedom. The mean square between litters in the analysis of variance for this experiment was 9.07 times as large as the error mean square ($s^2 = 41.192$), again demonstrating the importance of segregating litter-mate differences in growth assays for vitamin A.

TABLE XXXIII. Abbreviated Calculation of the Potency of Vitamin A β-Naphthoate from the Gain in Weight of Male Rats during a Five-Week Test Period. Collaborative Data Reported by Irwin (27). Log-dose interval $i = .3010$ or $S_2/S_1 = U_2/U_1 = 2$.

Litter No.	Growth on dose S_1	S_2	U_1	U_2	$D_1 = U_2 - S_2$	$D_2 = U_1 - S_1$	$D_3 = U_2 - U_1$	$D_4 = S_2 - S_1$	$y_a = D_1 + D_2$	$y_b = D_3 + D_4$	$y_{ab} = D_1 - D_2$
1	38	44	31	42	−2	−7	11	6	−9	17	5
2	57	62	50	64	2	−7	14	5	−5	19	9
3	29	45	31	52	7	2	21	16	9	37	5
4	23	38	16	34	−4	−7	18	15	−11	33	3
5	26	59	29	44	−15	3	15	33	−12	48	−18
6	32	40	24	47	7	−8	23	8	−1	31	15
7	43	51	30	40	−11	−13	10	8	−24	18	2
8	39	62	29	42	−20	−10	13	23	−30	36	−10
9	51	74	35	49	−25	−16	14	23	−41	37	−9
10	64	66	45	52	−14	−19	7	2	−33	9	5
T_d	402	541	320	466	−75	−82	146	139	−157	285	7
									$= T_a$	$= T_b$	$= T_{ab}$

$$s^2 = \frac{15041 - 105923/10}{12 \times 9} = 41.192 \quad \text{(Eq. 36)} \qquad R \begin{Bmatrix} 21 & 31 & 27 \\ 40 & 28 & 25 \end{Bmatrix}; \bar{R} = 28.667$$

$s = 6.4181, n = 27$ $\qquad s = .2150 \times 28.667 = 6.163$

The standard error of the T_i's as computed with the short-cut s was $2 \times 6.163 \times \sqrt{10} = 38.98$, leading by Eq. 28a to $t_b = 7.31$ and $t_{ab} = 0.18$, the one highly significant and the other nonsignificant. The potency was estimated by Eq. 29 as $M' = .3010\ (-157)/285 = -.1658$ and its error, s_M, by Eq. 32b. The slope $b = 285/(2 \times 10 \times .3010) = 47.34$ and $\lambda = 6.163/47.34 = .1302$, leading to

$$s_M = .1302 \sqrt{\frac{1}{10}\left(1 + \frac{157^2}{285^2 - (38.98 \times 2.052)^2}\right)} = .0475$$

Hence the relative potency was antilog $(-.1658) = .6827$ with a standard error of 11.6%. Since the standard solution contained 200 units per gram, the vitamin A content of the unknown was 137 ± 15 units per gram.

B. ASSAYS FROM PAIRED OBSERVATIONS

Experiments where a restriction in design should increase the precision of an assay may encounter a practical objection. In the vitamin A assay, for example, many litters will have only two or three of a given sex. These could not be used if the experiment were restricted to groups of four litter mates of the same sex. Or, if an animal is used repeatedly as in the thiamine assay, some individuals may die before four tests can be completed on each one. Moreover, the results are delayed since the fourth response must be completed before computing the potency of the unknown. Occasional gaps can be repaired by the missing-plot technique described in a preceding section but the losses may be too numerous to use this alternative.

The dilemma can be met in two ways. One is to use smaller groups and an incomplete balanced design in which every treatment occurs equally often with every other treatment within the same group. All treatment comparisons can then be recovered with equal precision. The application of this method to biological assays in groups of two or three is described elsewhere in detail and need not be discussed further here (33).

The other method is to set up the assay so as to "confound" one of the three treatment comparisons with groups. The "confounded" comparison is measured with less precision than the others, since its error is that between rather than within groups. In his discussion of the vitamin A assay, Gridgeman (34) has proposed confounding the slope. With this plan one-half of the available litter-mate pairs of a given sex are treated with the high dose of the standard and the high dose of the unknown and the remaining half with the low dose of the standard and the low dose of the unknown. The difference between the standard and the unknown (T_a) has an error associated with intra-litter comparisons whereas the slope (T_b) has an error based upon inter-litter comparisons. Two different error terms are required in estimating the precision of such an assay, the slope being determined with much less accuracy than the difference between the standard and the unknown. With this design the test for the parallelism of the two dosage-response curves (T_{ab}), however, has the same precision as the difference between standard and unknown (T_a). Since it does not enter directly into the computation of potency, the interaction of $a \times b$ is a more convenient and satisfactory comparison to confound than either of the others, as has been recommended for the insulin assay (34). T_{ab} may be confounded with group or pair differences by assigning one-half of the pairs to doses S_1 and U_2 and the other half to doses S_2 and U_1.

The results of the assay are analyzed in terms of the differences within pairs. The response on the low dose in each pair is subtracted from that on the high dose disregarding the distinction between standard and unknown. The intra-pair differences, $d_1 = U_2 - S_1$ and $d_2 = S_2 - U_1$, are then totaled for each type of pair to obtain the totals $S(d_1)$ and $S(d_2)$. The overall difference between standard and unknown is then computed as the difference between these two totals.

$$T_a = S(d_1) - S(d_2) = U_2 - S_1 - (S_2 - U_1) \tag{37}$$

The effect of slope is determined from the sum of the two totals or

$$T_b = S(d_1) + S(d_2) = U_2 - S_1 + S_2 - U_1 \tag{38}$$

Thus T_a and T_b have the same definition as in Table XXII and are used exactly as with complete groups of four in computing the log ratio of potencies M' by Eq. 29.

The error variance for T_a and T_b in units of a single response is given by the equation

$$s^2 = \frac{S(d^2)/2 - [S^2(d_1) + S^2(d_2)]/2f}{2(f-1)} \tag{39}$$

where f, as before, is the number of observations on each treatment. This estimate of the intra-pair error has $2(f-1)$ degrees of freedom. Alternatively and less efficiently, the standard deviation may be computed from the mean range $(\overline{R})$ in d_1 and in d_2, in this case multiplying $\overline{R}$ by the appropriate factor in Table XXXII. If the successive pairs are listed at random, they may be divided into equal groups of three to 9 values and the mean range $(\overline{R})$ determined from the d_i's within each subgroup. The standard error of the log ratio of potencies and the confidence limits in an experiment of this type may be computed from Eqs. 32a or 32b and by Eq. 34a as in other two-dose factorial assays.

The parallelism of the dosage-response curves for standard and unknown is tested with less precision than for complete groups from the sums of the responses in each pair, $p_1 = U_2 + S_1$ and $p_2 = S_2 + U_1$. The interaction of $a \times b$ which measures the departure from parallelism is computed from the totals $P_1 = S(p_1)$ and $P_2 = S(p_2)$ as

$$T_{ab} = P_1 - P_2 = S_1 + U_2 - (U_1 + S_2) \tag{40}$$

The inter-pair error for testing the significance of the difference in slope is based upon the squares of the pair totals or

$$s_p^2 = \frac{S(p^2)/2 - (P_1^2 + P_2^2)/2f}{2(f-1)} \tag{41}$$

TABLE XXXIV. Assay of Vitamin A from the Gain per Week of Male Rats during a Five-Week Period. Log Dose Interval = 0.6021 or $S_2/S_1 = U_2/U_1 = 4$.

Data from C. A. Morrell

Litter No.	Growth on S_2	Growth on U_1	$S_2 - U_1$ d_2	$S_2 + U_1$ p_2	Litter No.	Growth on U_2	Growth on S_1	$U_2 - S_1$ d_1	$U_2 + S_1$ p_1
1	4.6	0	4.6	4.6	10	3.7	0.8	2.9	4.5
2	3.4	2.1	1.3	5.5	11	3.0	2.6	0.4	5.6
3	7.8	4.7	3.1	12.5	12	6.4	5.3	1.1	11.7
4	9.4	4.9	4.5	14.3	13	7.7	4.7	3.0	12.4
5	8.6	2.9	5.7	11.5	14	6.0	5.3	0.7	11.3
6	5.6	1.7	3.9	7.3	15	7.1	5.2	1.9	12.3
7	5.3	2.7	2.6	8.0	16	7.1	5.4	1.7	12.5
8	5.5	0.9	4.6	6.4	17	4.1	3.3	0.8	7.4
9	8.4	—0.5	8.9	7.9	18	4.3	3.2	1.1	7.5
T_d	58.6	19.4	39.2	78.0	T_d	49.4	35.8	13.6	85.2
			$= S(d_2)$	$= P_2$				$= S(d_1)$	$= P_1$

$T_a = 13.6 - 39.2 = -25.6$ (Eq. 37)
$T_b = 13.6 + 39.2 = 52.8$ (Eq. 38)
} Standard error of T_a and $T_b = 1.171 \times 2\sqrt{9} = 7.026$

$M' = .6021(-25.6)/52.8 = -.2919$ (Eq. 29)

$T_{ab} = 85.2 - 78.0 = 7.2$ (Eq. 40) Standard error of $T_{ab} = 2.312 \times 2\sqrt{9} = 13.872$

$$s^2 = \frac{235.16/2 - 1721.60/18}{16} = \frac{21.9356}{16} = 1.3710;\ s = 1.171 \quad \text{(Eq. 39)}$$

$\overline{E}_d = (3.3 + 1.8 + 6.3 + 2.5 + 2.3 + 0.9)/6 = 2.850$; $s = .4178 \times 2.85 = 1.191$ from sets of 3.

$$s_p^2 = \frac{1653.56/2 - 741.28}{16} = 5.3438;\ s = 2.312 \quad \text{(Eq. 41)}$$

$\overline{E}_p = (7.9 + 7.0 + 1.6 + 7.2 + 1.1 + 5.1)/6 = 4.983$; $s = 2.082$ from sets of 3.

$b = 52.8/(2 \times .6021 \times 9) = 4.872$ (Eq. 7b)

$\lambda = 1.171/4.872 = .2404$ (Eq. 16)

$$s_M = .2404\sqrt{\frac{1}{9}\left(1 + \frac{25.6^2}{52.8^2 - (7.026 \times 2.120)^2}\right)} = .0898 \quad \text{(Eq. 32b)}$$

where p is the total of a single pair and P_1 and P_2 are their sums. The degrees of freedom are here equal to $2(f-1)$. In most cases where this design is applicable the F test for the significance of T_{ab} is of secondary interest, so that its lower precision is a minor handicap.

Example 12. The above procedure may be illustrated by the data in Table XXXIV from a five-week growth assay of vitamin A in male rats. The average growth per week was computed from body weights at the start of the test period and at the end of the third and fifth week on the test diet. The vitamin A was supplied in all cases by the International Standard. Doses of 2 and 8 units per day have been designated as S_1 and S_2 and doses of 1 and 4 units per day as U_1 and U_2. This gave a fourfold dosage-interval or $i = .6021$, and the true potency of the "unknown" was one-half that of the standard or the true $M = -.3010$. The average gain per week of each rat and the difference and sum of each litter-mate pair are shown in Table XXXIV. The essential calculations are given beneath the table.

The standard deviation estimated from the mean range agreed well with the efficient estimate in Eq. 39 for the intra-pair error but less closely for that between pairs. It is apparent from a comparison of $s = 1.171$ for intra-pair differences with $s = 2.312$ for inter-pair comparisons that the precision was increased materially by excluding the differences between litters from the determination of potency. The assayed potency of $M' = -.2919 \pm .0898$ agreed with the true value of $-.3010$ well within the sampling error.

C. ASSAYS WITH TWO UNKNOWNS IN GROUPS OF SIX

Factorial designs are readily extended to the assay of several unknowns in the same experiment. The procedure will be considered for the case

TABLE XXXV. Work Form for the Analysis of Variance of a Two-Dose Factorial Assay with Two Unknowns (U and U') and a Standard (S), Arranged in f Randomized Groups Each of Six Responses

	D.F.	Sum of squares	Mean square
Between groups	$f-1$	$S(T_g^2)/6 - C$	
Between preparations	2	$(T_s^2 + T_u^2 + T_{u'}^2)/2f - C$	
Combined slope	1	$B^2 = T_b^2/6f$	B^2
Nonparallelism	2	$[y^2]_b = S(B_t^2) - B^2$	
Error	$5(f-1)$	By difference	s^2
Total	$N-1$	$S(y^2) - C = [y^2]$	
Correction for mean	1	$C = S^2(y)/N$	

of two unknowns and one standard, each administered at two dosage levels which have the same ratio to each other in all preparations. The assay may be arranged in randomized groups of six or, if two restrictions are needed, in 6×6 Latin squares. The analysis will be considered in terms of randomized groups, although its extension to Latin squares involves no basic change in the calculation. In either case the assay slope of the combined dosage-response curve and the assay error are based upon the data for all three preparations, the two unknowns (U and U') and the standard (S).

The analysis of variance (Table XXXV) follows the usual form for randomized groups except in the subdivision of the 5 degrees of freedom among the six different doses or treatments. The sum of squares between preparations with 2 degrees of freedom is determined from the total of the responses on the standard and on each of the unknowns (T_s, T_u, and T_u') as shown in the work form. The mean square for this component tests whether the assayed potencies of the unknowns differ significantly from their assumed values. The effect of the combined slope is computed from T_b, which is equal to the total of all responses on the high doses minus the total of all responses on the low doses or

$$T_b = S(y_2) - S(y_1). \tag{42}$$

The test for nonparallelism is computed by an extension of Eq. 26 as indicated in Table XXXV, where each B_i^2 is computed separately from the T_{b_i} for each preparation as $B_i^2 = T_{b_i}^2/2f$. If the mean square for this row in the table is not significantly greater than the error, the potency of each preparation is calculated from the combined or assay slope. The total of the three sums of squares is equal exactly to the sum of squares for treatments computed as $S(T_a^2)/f - C$, which checks the arithmetic. The principal functions of the analysis of variance, therefore, are to measure the assay error s^2 and to determine whether the three preparations can be represented by a common slope which differs significantly from zero.

The potency of each unknown is computed by comparison with the same standard. Each log relative potency M' is determined as

$$M' = -\frac{3iT_a}{2T_b} \tag{43}$$

where T_a is computed from the difference in the response to a given unknown and that to the standard as defined in the first line of the work form in Table XXII and T_b is defined by Eq. 42. The combined slope of the assay is calculated by an extension of Eq. 7 as

$$b = \frac{T_b}{3if} \tag{7d}$$

where i is the log-dose interval between the two dosage levels of the standard and the unknown, and f is the number of observations at each dose of each preparation. The standard deviation from the mean square for error (s^2) is divided by this estimated slope to obtain λ for use in computing the standard error of M as

$$s_M = \lambda \sqrt{\frac{1}{f}\left(1 + \frac{3T_a^2}{2T_b^2 - 12fs^2i^2}\right)} \tag{32c}$$

This value of s_M may be used to compute the confidence limits as described in a preceding section.

Example 13. The calculation of the log potency and its standard error from an assay with two unknowns may be illustrated by the standardization of vitamin D from the line test in a routine experiment at the duPont Laboratories. The original data in Table XXXVI represent the scores for radius and ulna in 12 six-rat litters. The litter totals (T_g) in the right-hand column are used in segregating group differences. The treatment totals (T_d) have been combined to obtain for each preparation the sum of the scores at the two dosage levels and the difference of the scores at the high dose minus those at the low dose. The analysis of variance computed from these observations by the work form in Table

TABLE XXXVI. Assay of Two Preparations of Vitamin D Oil (U and U') against U.S.P. Reference Cod Liver Oil No. 2 (S) at Dosage Levels of 4 and 8 Units by the Line Test in Rats

Data from E.I. duPont de Nemours and Co., Inc.

Litter	Healing score for treatment						Total
No.	S_1	S_2	U_1	U_2	U_1'	U_2'	T_g
1	8	10	2	10	2	10	42
2	4	8	4	8	4	8	36
3	6	10	6	8	3	10	43
4	4	10	8	10	6	10	48
5	4	10	6	10	3	8	41
6	0	8	4	8	0	10	30
7	4	6	2	8	4	10	34
8	4	4	0	6	1	4	19
9	1	10	2	8	4	10	35
10	8	12	3	12	2	12	49
11	3	6	2	4	2	4	21
12	3	10	2	10	6	12	43
T_d	49	104	41	102	37	108	441
Sum	$T_s = 153$		$T_u = 143$		$T_{u'} = 145$		441
Difference	$T_{bs} = 55$		$T_{bu} = 61$		$T_{bu'} = 71$		187

XXXV is given in Table XXXVII. The small F value between oils indicates that the assumed potencies of one unit in 1.1546×10^{-6} g. of U and of one unit in 0.8060×10^{-6} g. of U' were consistent with their assayed potencies. The next two lines show that the dosage-response curves for all three preparations could be fitted by parallel lines.

TABLE XXXVII. Analysis of Variance of Vitamin D Assay in Table XXXVI

	D.F.	Sum of squares	Mean square	F
Between litters	11	170.05	15.459	4.99
Between oils	2	2.34	1.170	.38
Combined slope	1	485.68	485.68	156.77
Nonparallelism	2	5.44	2.720	.88
Error	55	170.37	3.098	1.00
Total	71	833.88		
Correction for mean	1	2701.12		

The log potency of each unknown relative to the value assumed in setting up the assay was computed from Eq. 43 as

$$M' = \frac{3 \times .3010\ (143 - 153)}{2 \times 187} = -.02414 \quad \text{for } U$$

and

$$M' = \frac{3 \times .3010\ (145 - 153)}{2 \times 187} = -.01932 \quad \text{for } U'$$

giving relative potencies of 0.9459 and 0.9565 in terms of their assumed values. In units of vitamin D per gram, the first unknown (U) assayed at $.9459 \times 10^6/1.1546 = 819{,}000$ units per gram and the second unknown at 1,187,000 units per gram. The combined slope by Eq. 7d was $b = 187/(3 \times .3010 \times 12) = 17.257$, from which $\lambda = \sqrt{3.098}/17.257 = 0.1020$. Equation 32c was then solved with the data for U to obtain

$$s_M = .1020 \sqrt{\frac{1}{12}\left(1 + \frac{3 \times 10^2}{2 \times 187^2 - 12 \times 12 \times 3.098 \times 2.006^2}\right)} = .02951$$

from which the value for U' did not differ appreciably. The antilogarithm of s_M indicates a standard error of 7.0% of the asayed potency.

D. ASSAYS WITH TWO UNKNOWNS IN GROUPS OF THREE

The availability of the two-dose factorial assays described in the preceding section may be limited by the small size of homogeneous groups of assay animals. Thus litters containing six rats or more of a given sex form only a small percentage of those born. When two unknowns are ready for testing at the same time, the two-dose assay may be used

with litters of three by confounding the test for parallelism with differences between litters. The unknowns in such multiple assays should be similar, so as to minimize possible differences from the Standard in slope. The technique will be described for randomized groups of three although the design can be arranged in Latin squares when it is desirable to segregate two identifiable sources of variation. It can also be extended to three or more unknowns if required.

The simultaneous assay of two unknowns against the Reference Standard requires groups of three responses in two sets of equal size. The animals in each group are assigned at random to the treatments designated by the letters. In set I each group tests the low dose of the Standard (S_1) and the high dose of both unknowns (U_2 and U_2'), in set II each group tests the high dose of the Standard (S_2) and the low dose of both unknowns (U_1 and U_1'). In a vitamin A assay, for example, the groups might be litters of three male rats with half of them assigned to set I and half to set II.

Perhaps the simplest analysis is that based upon two differences (y_h and y_j) and a sum (y_k) computed from the three responses in each group of both sets. These are defined as

$$\begin{aligned} y_h &= y_u - y_{u'} \\ y_j &= y_u + y_{u'} - 2y_s \\ y_k &= y_u + y_{u'} + 2y_s \end{aligned} \tag{44}$$

where the subscripts on the right refer to the two unknowns and the standard. The individual responses and the differences and sum for each group are totaled separately in each set. The assay slope is based upon the data for all three preparations and depends upon the difference in y_j between sets, giving

$$T_b = S(y_j)_{\text{I}} - S(y_j)_{\text{II}} \tag{42a}$$

In terms of its assumed vitamin content, the log relative potency of each unknown in the presence of confounding is determined as

$$M' = \frac{3iT_a}{T_b} \tag{43a}$$

where i is the log interval between dosage levels and T_a is defined (Table XXII) as $T_a = S(y_u) - S(y_s)$ for one unknown and as $S(y_{u'}) - S(y_s)$ for the other unknown.

The assay slope reduces in the present case to

$$b = \frac{T_b}{6if} \tag{7e}$$

where f is the number of responses on each of the six different dosage levels or treatments. To validate its use, two degrees of freedom are available for testing whether the individual dosage-response curves for the three preparations are parallel within the experimental error. These, in turn, depend upon two different error variances. The error variance within litters s^2, which is also used in computing the standard error of the assayed potency, depends upon the sum of squares

$$[y_h^2] = \frac{S(y_h^2)}{2} - \frac{S^2(y_h)_{\mathrm{I}} + S^2(y_h)_{\mathrm{II}}}{2f}$$

and (45)

$$[y_j^2] = \frac{S(y_j^2)}{6} - \frac{S^2(y_j)_{\mathrm{I}} + S^2(y_j)_{\mathrm{II}}}{6f}$$

from which

$$s^2 = \frac{[y_h^2] + [y_j^2]}{4(f-1)} \tag{46}$$

The variance ratio testing the divergence of the dosage-response curves for the two unknowns is then determined as

$$F_h = \frac{\{S(y_h)_{\mathrm{I}} - S(y_h)_{\mathrm{II}}\}^2}{4fs^2} \tag{47}$$

for which $n_1 = 1$ and n_2 is the degrees of freedom in s^2 or $4(f-1)$ in a complete experiment. The divergence in the slope of the combined curve for the two unknowns from that for the standard depends upon a comparison between rather than within litters. The appropriate group variance is determined from the sum of squares

$$[y_k^2] = \frac{S(y_k^2)}{6} - \frac{S^2(y_k)_{\mathrm{I}} + S^2(y_k)_{\mathrm{II}}}{6f} \tag{45a}$$

as

$$s_k^2 = \frac{[y_k^2]}{2(f-1)} \tag{46a}$$

The variance ratio for testing the discrepancy in slope between standard and the unknown is then

$$F_k = \frac{\{S(y_k)_{\mathrm{I}} - S(y_k)_{\mathrm{II}}\}^2}{12fs_k^2} \tag{47a}$$

for which $n_1 = 1$ and $n_2 = 2(f-1)$. If neither F_h nor F_k is significantly larger than unity, the combined assay slope may be used in computing the potency of each unknown. If the slopes diverge, the calculation of potency should be restricted to those preparations for which the slopes are parallel.

TABLE XXXVIII. Two-Dose Factorial Assay of Two Preparations of Vitamin A (U and U') against U.S.P. Vitamin A Reference Standard (S) with Two Sets (I and II) Each of Eight Three-Rat Litters of Males; Growth Measured in Grams per Week over a Four-Week Test Period. Differences y_h and y_j and Sums y_k Defined in Eqs. 44.

Data from Laboratory of Vitamin Technology

Litter	Growth in set I on						Litter	Growth in set II on					
No.	U_2	U_2'	S_1	y_h	y_j	y_k	No.	U_1	U_1'	S_2	y_h	y_j	y_k
1	17.4	23.4	18.7	— 6.0	3.4	78.2	2	14.3	13.4[a]	19.9	.9	— 12.1	67.5
3	22.2	22.4	20.7	— .2	3.2	86.0	4	14.1	11.9	23.1	2.2	— 20.2	72.2
5	19.0	18.3	14.5	.7	8.3	66.3	6	8.5	13.6	16.4	— 5.1	— 10.7	54.9
7	19.0	22.1	15.8	— 3.1	9.5	72.7	8	13.5	12.2	21.1	1.3	— 16.5	67.9
9	16.7	14.1	13.1	2.6	4.6	57.0	10	6.2	10.2	20.8	— 4.0	— 25.2	58.0
11	19.1	28.1	17.7	— 9.0	11.8	82.6	12	10.3	15.5	27.1	— 5.2	— 28.4	80.0
13	15.1	21.1	14.5	— 6.0	7.2	65.2	14	11.7	17.8	24.8	— 6.1	— 20.1	79.1
15	18.8	21.7	13.7	— 2.9	13.1	67.9	16	8.1	8.1	24.0	0	— 31.8	64.2
T_d	147.3	171.2	128.7	— 23.9	61.1	575.9	T_d	86.7	102.7	177.2	— 16.0	— 165.0	543.8

[a] Replacement for missing value as determined by Eq. 21.

$$T_b = 61.1 - (-165.0) = 226.1 \quad \text{(Eq. 42a)}, \qquad b = 226.1/(6 \times .3010 \times 8) = 15.649 \quad \text{(Eq. 7e)}$$

$$T_a = 147.3 - 128.7 + 86.7 - 177.2 = -71.9, \qquad M'_u = 3(.301)(-71.9)/226.1 = -.2872 \quad \text{(Eq. 43a)}$$

$$T_a' = 171.2 - 128.7 + 102.7 - 177.2 = -32.0, \qquad M'_{u'} = 3(.301)(-32.0)/226.1 = -.1278 \quad \text{(Eq. 43a)}$$

$$[y_h^2] = 291.9/2 - (23.9^2 + 16.0^2)/(2 \times 8) = 94.25$$
$$[y_j^2] = 4362.83/6 - (61.1^2 + 165.0^2)/(6 \times 8) = 82.18 \quad \text{(Eq. 45)}$$

$$[y_k^2] = 79661.39/6 - (575.9^2 + 543.8^2)/(6 \times 8) = 206.50 \quad \text{(Eq. 45a)}$$

$$s^2 = (94.25 + 82.18)/(4 \times 7 - 1) = 6.534, \; n = 27; \quad s = 2.556 \quad \text{(Eq. 46)}$$

$$s_k^2 = 206.50/(2 \times 7) = 14.750, \; n = 14; \quad s_k = 3.841 \quad \text{(Eq. 46a)}$$

$$F_h = (-23.9 + 16.0)^2/(4 \times 8 \times 6.534) = 0.30 \quad \text{(Eq. 47)}$$

$$F_k = (575.9 - 543.8)^2/(12 \times 8 \times 14.750) = 0.73 \quad \text{(Eq. 47a)}$$

$$\lambda = 2.556/15.649 = .1633; \quad t = 2.052 \text{ at } P = .05 \text{ and } n = 27$$

$$s_M = .1633 \sqrt{\frac{1}{8}\left(1 + \frac{3T_a^2}{226.1^2 - 12 \times 8 \times 6.534 \times 2.052^2}\right)} = .0663 \text{ for } U \text{ and } .0595 \text{ for } U' \quad \text{(Eq. 32d)}$$

Assuming that both log potencies can be computed with the same assay slope, the standard error of M may be derived as before to obtain

$$s_M = \lambda \sqrt{\frac{1}{f}\left(1 + \frac{3T_a^2}{T_b^2 - 12fs^2t^2}\right)} \tag{32d}$$

where λ is computed with the standard deviation (s) within groups as determined by Eq. 46, and t is the value expected at a given level of significance for the degrees of freedom in s^2. If the experimenter is willing to forfeit some efficiency in his estimates, s and s_k may be computed from the mean range in y_h, y_j, and y_k. When f exceeds 10 or 11, the sets should be subdivided, preferably into equal subsets, and the range determined in each subset. These ranges are then averaged to obtain $\overline{R}_h$, $\overline{R}_j$, and $\overline{R}_k$, respectively. The intra-group standard deviation is estimated approximately as

$$s = \frac{K}{2}\left(\overline{R}_h + \frac{\overline{R}_j}{\sqrt{3}}\right) \tag{48}$$

where K is the coefficient in the last column of Table XXXII corresponding to the number of entries in each set or subset used in determining the R's. The inter-group error for testing parallelism may be estimated similarly as

$$s_k = \frac{K\overline{R}_k}{\sqrt{3}} \tag{48a}$$

If subsets vary in size, each R is multiplied by the appropriate K and the products (KR) are averaged in computing s or s_k.

Example 14. The two-dose factorial assay of two unknowns with groups of three may be illustrated by a special growth test on vitamin A with male rats as reported by the Laboratory of Vitamin Technology. The average growth rate in grams per week over a four-week test period is given for each rat in Table XXXVIII. The dosage levels were 1 and 2 observed or assumed U.S.P. units of vitamin A per day but due to improvements in the diet, the rate of gain exceeded that usually obtained at these dosages. One value (for U_1' in litter 2) was missing and has been replaced by applying Eq. 21 to the remaining data in set II. As a result there was one less degree of freedom in s^2.

After completing the record the first step was to compute y_h, y_j, and y_k for each litter. The same procedure was applied to the totals T_d for an arithmetic check on the calculation of the y's. The calculation of the log potencies is shown beneath Table XXXVIII, leading to $M' = -0.2872$ for U and to -0.1278 for U' or assayed potencies of 51.6 and 74.5% of their assumed values.

These values were determined with the composite assay slope. The test of its validity required both the variance within litters and between litters as defined by Eq. 46 and 46a. The variance within litters, $s^2 = 6.534$ with 27 degrees of freedom was appreciably smaller than that between litters, $s_k^2 = 14.750$ with 14 degrees of freedom, demonstrating again the desirability of segregating litter differences in the vitamin A assay. From the variance ratio testing the divergence in slope between the two unknowns ($F_h = 0.30$) and that testing their average difference from the standard ($F_k = 0.73$), the three contributions to the assay slope could be considered as parallel.

By virtue of the design the standard error of potency involved only the variation between rats within litters. The ratio of the standard deviation to slope was $\lambda = 0.1633$, from which $s_M = 0.0663$ for U and 0.0595 for U' when computed by Eq. 32d as shown beneath the table. The standard deviation within litters has also been computed from the range of the eight values of y_h and of y_j in sets I and II, giving $\overline{R}_h = (11.6 + 8.3)/2 = 9.95$ and $\overline{R}_j = (9.9 + 21.1)/2 = 15.5$. Substituting these in Eq. 48, $s = .2484\ (9.95 + 15.5/1.732)/2 = 2.35$, which may be compared with the efficient estimate by Eq. 46 of $s = 2.556$. The approximate estimate of $s_k = 3.88$ by Eq. 48a may be compared similarly with $s_k = 3.841$ by Eq. 46a.

V. The Correction of Quantitative Variables: Covariance

A number of quantitative variables may affect the precision of vitamin assays which cannot be segregated readily by restrictions in the design. Examples are the weight and age at depletion of rats in the growth assay for vitamin A. The U.S. Pharmacopoeia specifies that individuals are to be assigned to different treatments so as to minimize differences in weight and in the number of rats starting the test period on any one day. This requirement assumes that both the weight and time of depletion may affect rat growth during the test period. While factors such as these often can be balanced by assigning animals to doses with the aid of a Latin square, computation as a Latin square does not measure satisfactorily their effect upon the precision of an assay. This is the purpose of covariance.

Covariance will show whether a suspected quantitative factor has a real effect and then will correct it efficiently. Vitamin dosages in animal assays, for example, are commonly given at a constant amount per animal. Yet the effect produced by many drugs depends upon the size of the animal, so that doses are expressed in milligrams of drug per kilogram of body weight. This correction in turn has proved inaccurate in several

cases and cannot be recommended blindly. By covariance, however, it is possible to determine an adjustment for body weight which will minimize the error in a given assay, as has been done by Gridgeman in a line test for vitamin D (36).

Sometimes an arbitrary correction may be concealed in the response used for measuring potency. The percentage bone ash in the rat assay for vitamin D, for example, involves such an untested assumption (37). The original measurements consist of the weight of the fat-extracted bone and its weight after ashing. Their combination into a percentage ash implies that the logarithm of the ash content increases proportionately with the logarithm of the organic content lost in ashing. If the lines relating these two variables within each dosage level were to differ appreciably in slope from unity or from linearity, the percentage ash would not be an efficient criterion for measuring vitamin D. This has proved to be the case in actual test (36, 37). Thus covariance is also of value in testing the validity of hidden assumptions and in selecting a more efficient criterion for measuring vitamin potency.

Covariance finds its greatest potential value in improving assay procedure. If the usual response were to show the same quantitative relation to a concomitant measure in successive experiments, the assay criterion could be redefined so as to include the adjustment without recomputing it each time. If a factor should prove consistently of no importance in typical assays, it could be neglected in later experiments without loss of precision. Thus in computing the results of several growth assays of vitamin A in which differences between litters could be segregated, the age and weight of the rat at depletion were found to be completely negligible when tested by covariance. If this finding were confirmed, several of the restrictions in the U.S.P. XIV assay for vitamin A could be relaxed without loss of precision.

1. THE COVARIANCE BETWEEN A CONCOMITANT MEASURE AND THE RESPONSE

Covariance is based upon a relatively simple application of linear regression. In effect, the response is plotted against a concomitant measurement which is unaffected by the dosage of vitamin but is suspected of modifying the response. The first variate is related to the second by a straight line which is fitted by least squares. If the slope of this line differs significantly from zero, the relationship is presumably real and can be used to reduce the variation in the response. The slope is computed so that it is independent of differences between doses or between restrictions in design. It represents, in effect, the best-fitting series of parallel straight lines within dosage levels and within randomized groups

or their equivalent. If the relation is not linear, a curved regression may be used instead, but the computation is then considerably more tedious. Whenever possible this is avoided by converting the regression to a linear form by a suitable transformation of the concomitant measure. Two or more concomitant measures can be allowed for by covariance but such cases are beyond the scope of this chapter.

The computation of covariance is a simple extension of the analysis of variance described in the preceding sections. In the analysis of variance the total sum of the squared deviations between each individual response (y) and the general mean ($\bar{y}$) is separated into its relevant components. Although uninfluenced by differences in dosage, the total variation in the initial or concomitant variable (v) can be subdivided into the same components to form a parallel table of sums of squares. The calculation of the two series of sums of squares is identical except that the $[v^2]$'s are computed from the concomitant variate and the $[y^2]$'s from the response. Both are required.

A third set of values, the sums of the products or the $[vy]$'s, completes the basic computation. In each case a value within square brackets, [ ], represents a sum of the squares or of the products of deviations from their respective means. The products $[vy]$ can be calculated by following a very simple rule: at each and every stage where a number would be squared in calculating $[v^2]$ and $[y^2]$, the corresponding values of v and y are multiplied together to obtain $[vy]$. Unlike the sums of squares, which must always be positive, the sums of products may be either positive or negative in sign.

In the completed table all entries in the row for error are unaffected by variation attributable to differences between the restrictions in design or between the dosages of vitamin. Hence the straight line computed from the entries in the row for error provides an unbiased estimate of the relation of y to v. Its slope is determined by Eq. 6 as $b_v = [vy]/[v^2]$ and can be used to adjust y for variation in v. If y is related to v so that y decreases as v increases, the slope b will be negative.

The first objective in the analysis of covariance is to determine whether b_v differs more from a slope of zero than its sampling error. Unless it does, y is presumably unaffected by variations in v in a given experiment. To measure the effect of the regression in reducing the variance in y, B_v^2 is computed by Eq. 9 from the numerator and denominator of b_v. B_v^2 is then subtracted from $[y^2]$ in the same row for error and the resulting "reduced $[y^2]$" has one less degree of freedom than before. Its mean square is the error variance s^2 with which the significance of B_v^2 may be tested from the variance ratio $F = B_v^2/s^2$.

In applying the method of covariance, the treatment sum of squares for the concomitant measurement should not be significantly larger than its experimental error. If the concomitant measure were modified significantly by the experimental treatments, it might account for part of the response and we could lose a substantial part of the effect of the vitamin in "correcting" for the supposedly concomitant value. In some cases this loss may be only apparent but usually one prefers to avoid it. For this reason the log weight of the organic content (= loss of weight on ignition) has been used as the concomitant variable in the following example, rather than the log weight of the fat-extracted bone as has been proposed by Gridgeman (36). Where the dependent variate forms a substantial part of the concomitant measure, the interpretation of the analysis of covariance is much less direct.

Example 15a. The method may be applied to data reported by Coward (11) on the standardization of vitamin D from the ash content of the femur of the rat. Six litter mates were used from each of six litters. In each litter, three rats received daily, over a period of six weeks, 0.025, 0.05, and 0.1 unit of international standard vitamin D and the other three an unknown cod liver oil at rates of 0.2, 0.4, and 0.8 mg. daily for the same period. At the end of the test the femora were removed from each rat, fat extracted with alcohol, dried to constant weight and ashed. Table XXXIX shows for each individual the log weight of the organic content (v) and of the ash (y) (37).

TABLE XXXIX. Organic and Ash Content of Dry Bone Observed in an Assay of Vitamin D in Rats

Data of Coward (11, 37)

	Litter No.	Data for treatment						Total
		S_1	S_2	S_3	U_1	U_2	U_3	T_g
Log weight of organic content (—.7) (v)	1	.092	.014	.193	.176	.101	.125	.701
	2	.120	.118	.092	.090	.133	.129	.682
	3	.177	.216	.193	.104	.225	.228	1.143
	4	.105	.053	.146	.087	.134	.185	.710
	5	.102	.109	.111	.211	.019	.099	.651
	6	.093	.074	.165	.083	.131	.186	.732
	T_d (=V_d)	.689	.584	.900	.751	.743	.952	4.619
Log weight of ash (y)	1	.207	.494	.633	.396	.537	.702	2.969
	2	.346	.459	.640	.467	.590	.760	3.262
	3	.377	.606	.751	.511	.678	.880	3.803
	4	.281	.450	.720	.449	.619	.820	3.339
	5	.318	.431	.674	.423	.464	.778	3.088
	6	.262	.453	.708	.498	.614	.778	3.313
	T_d (=Y_d)	1.791	2.893	4.126	2.744	3.502	4.718	19.774

$S(v^2) = 0.694773$; $S(vy) = 2.684504$; $S(y^2) = 11.881014$

The total sums of squares and products have been divided into three parts by the analysis of covariance in Table XL. These represent the differences between litters or rows, the differences between treatments or columns, and the interaction of treatments × litters or the experimental error. Differences between treatments will later be subdivided factorially

TABLE XL. Analysis of Covariance for the Vitamin D Assay in Table XXXIX

		Sums of squares and products		
Variation due to	D.F.	$[v^2]$	$[vy]$	$[y^2]$
Litters	5	.02847	.03936	.06842
Doses of vitamin D	5	.01538	.09475	.91426
Experimental error	25	.05828	.01328	.03691
Total	35	.10213	.14739	1.01959
Correction for mean	1	.59264	2.53711	10.86142

From the row for error:

$$b_v = \frac{.01328}{.05828} = .2278 \quad \text{(Eq. 6)} \qquad B_v{}^2 = \frac{.01328^2}{.05828} = .003026 \quad \text{(Eq. 9)}$$

$$\text{Reduced error variance, } s^2 = \frac{.03691 - .00303}{25 - 1} = .001412 \quad .$$

but at this stage it is more convenient to handle them together. With this exception the computation of the sums of squares followed the work form for randomized groups in Table X. The sum of squares for doses in the column for $[v^2]$, for example, was computed as $S(V_d{}^2)/k - C = (.689^2 + .584^2 + \cdots + .952^2)/6 - .59264 = .01538$. The sums of products were obtained in the same manner as the sums of squares except that corresponding values of v and y were cross-multiplied without other change in the equations. Thus the correction term for the sums of products was $(4.619 \times 19.774)/36 = 2.53711$. The sum of products between litters was computed as $(.701 \times 2.969 + .682 \times 3.262 + \cdots + .732 \times 3.313)/6 - 2.53711 = .03936$ and that for doses similarly.

The regression of y on v within doses and litters was computed from the row for experimental error to obtain $b_v = .2278$. If the percentage ash were a suitable measure of the response to vitamin D, a slope of $b_v = 1$ would be expected (37). In this experiment, at least, the observed slope was significantly smaller ($P < .001$) than that assumed in using the percentage ash. The variance accounted for by the slope was $B_v{}^2 = .003026$. This was compared with the reduced error variance, $s^2 = .001412$, to obtain $F = .003026/.001412 = 2.14$ with $n_1 = 1, n_2 = 24$, and $0.1 < P < 0.2$. Hence the slope of y upon v, $b_v = .2278$, did not differ significantly from zero.

At this stage one could conclude that variations in the organic content of the bone had no relevance in this assay of vitamin D and compute the log ratio of potency from the log weight of ash alone. However, the regression of y upon v exceeded its error, even though not significantly, so that the experiment will be used to exemplify the calculation in the next section. In terms of v the variance ratio for the average effect of dose was $F = 1.32$, in marked contrast to that for y, where $F = 123.8$. Since differences in the dose of vitamin D apparently had no effect upon the organic content, v can serve as a concomitant measure without jeopardizing the response to the assay.

2. The Adjusted Estimate of Potency

Given the net relation of the response (y) to the initial or concomitant measure (v), it would be possible to adjust each observation as

$$y' = y - b_v(v - \bar{v})$$

Each adjusted response, y', would then have the value expected if its corresponding v were equal exactly to $\bar{v}$ or the same for all animals in the test. A separate analysis of variance might then be computed from the adjusted y's. The same result, however, can be obtained more easily by adjusting the treatment effects directly for v.

In determining the significance of v by an analysis of covariance, it may be convenient to pool all treatment effects initially into a single term. In order to estimate the potency of the unknown, however, they must be separated into the components described in the preceding sections. If the experiment has been designed factorially, a scheme such as that for the two-dose or the three-dose assay in Tables XXII or XXIV is available. The work forms are modified by listing two rows of dose totals (T_d) beneath the columns of factorial coefficients, V_d for v and Y_d for y respectively. These dose totals are multiplied in turn by the factorial coefficients to obtain two sums of products[2] $S(xV_d)$ and $S(xY_d)$ for each treatment effect. If the sums of products in any given row i are designated as T_{iv} and T_{iy} for the concomitant measure and the response respectively, each treatment effect is adjusted for differences in v by computing

$$T_i' = T_{iy} - b_v T_{iv} \tag{49}$$

[2] Note that the term "sum of products" is used in this section to refer both to $[vy]$ the sum of products of deviations from their respective means, as in Table XL and to $S(xV_d)$ or $S(xY_d)$, the sum of the products of factorial coefficients by the dose totals, as in Table XLI.

TABLE XLI. Factorial Analysis of the Vitamin D Assay in Table XXXIX with the Coefficients in Table XXIV

Variance due to	Factorial coefficients (x) for dose						Divisor	Sum of products		Adjusted treatment effects
	S_1	S_2	S_3	U_1	U_2	U_3	$fS(x^2)$	$S(xV_d)$	$S(xY_d)$	
(a) Difference in S and U	—1	—1	—1	1	1	1	36	.273	2.154	$2.092 = T_a'$
(b) Combined slope	—1	0	1	—1	0	1	24	.412	4.309	$4.215 = T_b'$
(ab) Lines not parallel	1	0	—1	—1	0	1	24	—.010	—.361	$-.359 = T_{ab}'$
(o) Similar curvature	1	—2	1	1	—2	1	72	.638	.589	$.444 = T_o'$
(ac) Opposed curvature	—1	2	—1	1	—2	1	72	—.204	.327	$.373 = T_{ao}'$
T_d for $v\ (=V_d)$	.689	.584	.900	.751	.743	.952				$b_v = .2278$
$y\ (=Y_d)$	1.791	2.893	4.126	2.744	3.502	4.718				

$$i = .3010 \qquad M' = \frac{4 \times .3010 \times 2.092}{3 \times 4.215} = .1992 \quad \text{(Eq. 30)}$$

Thus the adjusted treatment effects in the first two rows are $T_a' = T_{ay} - b_v T_{av}$ and $T_b' = T_{by} - b_v T_{bv}$. The log ratio of potencies M', adjusted for variations in v, is obtained by substituting T_a' and T_b' in Eq. 29 for a two-dose assay and in Eq. 30 for a three-dose assay. A comparable adjustment is available for nonfactorial assays.

The significance of each adjusted T_i' can be tested approximately with Eq. 28 by computing the statistic t with the adjusted value of T_i and of s^2. The resulting t tends to overestimate the significance of each factor since it does not allow for the error in b_v, the slope used in making the adjustment. Wherever the approximate t indicates no significance, an unbiased test will also show no significance. Factors which are barely significant by the approximate t may not be with an unbiased test but those where the significance is marked are seldom changed.

Example 15b. The five treatment effects in the vitamin D assay of Table XXXIX have been isolated factorially in Table XLI. The dose totals (T_d) for the concomitant measure V_d and for the response Y_d from Table XXXIX have been multiplied in turn by the coefficients in each row and summed to obtain the two columns of sums of products, $S(xV_d)$ and $S(xY_d)$. Each treatment effect was then adjusted for differences in v with the slope $b_v = .2278$ from Table XL by substitution in Eq. 49. The first row, for example, gave $T_a' = 2.154 - .2278 \times .273 = 2.092$ and the others were obtained similarly.

The validity of the assay could be checked by the approximate test for the significance of each adjusted T_i'. The reduced error variance from Table XL, $s^2 = .001412$, was multiplied in turn by each divisor and an approximate t computed by Eq. 28, that for T_b', for example, being approximate $t = 4.215/\sqrt{.001412 \times 24} = 4.215/.184 = 22.9$, which was very highly significant as is characteristic of a good assay. Although larger than their respective errors, none of the last three factors, T_{ab}', T_c' and T_{ac}', was significant at $P = .05$, their approximate t values being 1.95, 1.39, and 1.17 respectively with 24 degrees of freedom. Hence the assay would be considered acceptable.

The relative potency of the unknown cod liver oil was computed by substituting T_a' and T_b' in Eq. 30, the dosage interval being $i = \log 2 = .3010$. Then $M' = .1992$ in logarithmic units. From the antilogarithm of .1992, the potency of the unknown in units assumed to be equivalent to the standard was 158.2%. Since each gram of the cod liver oil was assumed to contain 125 International Units, the potency of the unknown has been assayed as $1.582 \times 125 = 197.8$ units per gram. For comparison, the log relative potency computed from the unadjusted $T_a = 2.154$ and $T_b = 4.309$ was $M' = 0.2006$, leading to an assayed potency of 198.4 units per gram.

3. The Error of the Estimated Potency

The log potency is easy to compute with the adjusted $(\bar{y}_u - \bar{y}_s)$ and slope or with T_a' and T_b' and, as described above, leads to an unbiased adjustment for the concomitant measurement. Its exact standard error and confidence limits, however, are considerably more difficult to compute than in the absence of covariance. As Finney (38) has pointed out, there will be sampling variation from zero correlation between the dose of vitamin (x) and the concomitant variate (v). This does not affect the log potency as computed from the adjusted T_a' and T_b' or their equivalent but s_M is modified by a covariance term between x and v. This requires the computation of the standard errors of partial regression coefficients, which is beyond the scope of this chapter. For an exact solution the reader is referred to Finney's paper (38).

Fortunately, an approximation, which is adequate for most purposes, can be computed with the adjusted treatment effects T_a' and T_b' or their equivalent and the corrected error variance s^2. These are used in determining the standard error s_M by Eq. 32 or 32a and the confidence limits X_L by Eq. 34 or 34a. D^2 and B^2 are estimated for this purpose as $T_a'^2/fS(x_a^2)$ and $T_b'^2/fS(x_b^2)$. Both values are uncorrected for the sampling variation in b_v, the slope of the response plotted against the concomitant measure, but this discrepancy may be small relative to that caused by omitting the covariance of x and v. For most purposes the discrepancy in using the adjusted values may be overlooked.

Where the adjustment for v is substantial, the exact standard error and confidence limits may be approximated more closely by using the so-called "reduced" variances for D^2 and B^2 (3, 39). These are also needed for critical tests of significance. The corrected error variance is computed from a "reduced" sum of squares in that the slope for correcting v is calculated from $[v^2]$ and $[vy]$ in the same row to which it is applied. Each of the other reduced $[y^2]$'s is computed similarly with the slope determined from the sums of squares and products, $[v^2]$, $[vy]$, and $[y^2]$, for a given factor added to the equivalent values in the row for experimental error. The effect of slope for the combination in each such supplementary row is $B_v^2 = [vy]^2/[v^2]$, which, in turn, is subtracted from $[y^2]$ in the same row to obtain the reduced $[y^2]$ with one less degree of freedom. But this value represents both the error and the given treatment effect, so that to recover the reduced $[y^2]$ for the treatment factor alone, the reduced $[y^2]$ based only on the error row must be subtracted. Each reduced treatment sum of squares, therefore, is obtained indirectly by subtracting the reduced $[y^2]$ for error alone from that for

the given treatment plus error. This process corrects automatically any error of estimate in equalizing the initial or concomitant measure v.

The reduced $[y^2]$'s needed in computing s_M are those for the difference in the response to the standard and to the unknown (D^2) and for the combined slope of the dosage-response curve (B^2), each with one degree of freedom. Two variances and a covariance are computed from the sums of products in each of the first two rows of the factorial analysis as

$$\begin{aligned} [v^2] &= S^2(xV_d)/fS(x^2) \\ [vy] &= S(xV_d)S(xY_d)/fS(x^2) \end{aligned} \tag{50}$$

and

$$[y^2] = S^2(xY_d)/fS(x^2)$$

Unless the adjusted values for the remaining treatment effects have thrown suspicion on the assay, their sums of squares and products need not be isolated individually. The $[v^2]$, $[vy]$, and $[y^2]$ for the first two treatment effects (a) and (b) are subtracted from the corresponding terms in the row for dose or treatments in the initial analysis of covariance. The reduced mean square for all remaining treatment effects should not be significant in a valid assay. The approximate standard error and confidence limits for M' are computed with the reduced values of D^2 and B^2 and with M' and λ as determined from the adjusted values of T_a' and T_b'.

Example 15c. The calculation of the approximate standard error of the log ratio of potencies may be illustrated with the vitamin D assay of Table XXXIX. From the second adjusted treatment effect in Table XLI, the combined slope of the dosage-response curve has been computed by Eq. 7c as $b = 4.215/(4 \times .3010 \times 6) = .5835$. The standard deviation in y from the reduced error variance in Table XL was $s = \sqrt{.001412} = .03758$, so that the standard deviation in units of x was $\lambda = .03758/.5835 = .06441$. Adjusted variances for the first two treatment comparisons were determined from T_a' and T_b' in Table XLI as $D^2 = 2.092^2/36 = .1216$ and $B^2 = 4.215^2/24 = .7403$. With $t = 2.064$ for $n = 24$ and $P = .05$, $B^2 - s^2t^2 = .7342$, leading by Eq. 32a to

$$s_M = .06441 \sqrt{\frac{4}{36}\left(1 + \frac{.1216}{.7342}\right)} = .02318$$

Since the antilogarithm of s_M was 1.055, the assayed potency of 197.8 units per gram had a standard error of 5.5%. The standard error could also be converted by Eq. 33 to units of vitamin D as $2.303 \times .02318 \times 197.8 = 10.6$ and the result of the assay reported as 197.8 ± 10.6 units per gram. Because the concomitant measure v did not have a significant

TABLE XLII. Computation of Reduced Treatment Variances from the Data in Tables XL and XLI

Variance due to	D.F.	Sums of squares and products $[v^2]$	$[vy]$	$[y^2]$	B_v^2	$[y^2]$ Reduced for v: D.F.	Sum of squares	Mean square	F
(*a*) Standard vs. Unknown	1	.00207	.01633	.12888		1	.11738	$.11738 = D^2$	83.13
(*b*) Combined slope	1	.00707	.07397	.77364		1	.66018	$.66018 = B^2$	467.5
Other treatment effects	3	.00624	.00444	.01173		3	.00989	.003297	2.33
Error	25	.05828	.01328	.03691	.00303	24	.03388	$.001412 = s^2$	1.00
Error + S vs. U	26	.06035	.02961	.16579	.01453	25	.15126		
Error + slope	26	.06535	.08725	.81055	.11649	25	.69406		
Error + other treatments	28	.06452	.01772	.04864	.00487	27	.04377		

$B^2 - s^2t^2 = .6542$ $\quad C^2 = 1.0091$ (Eq. 35) $\quad C = 1.0045$ $\quad t = 2.064$

$s_M = .02332$ (Eq. 32a) $\quad X_L = .2010 \pm .0483 = .1527$ and $.2493$

effect upon the response y, the precision of the unadjusted value should be nearly the same. Without covariance $s_M = .02321$.

The reduced $[y^2]$'s are not needed in the present example but have been computed in Table XLII to illustrate the method. The variances and covariance for (a), the standard vs. the unknown, were determined from the sums of products in the first row of Table XLI as $.273^2/36 = .00207$, $.273 \times 2.154/36 = .01633$ and $2.154^2/36 = .12888$ respectively. The corresponding terms for (b), the combined slope, were computed similarly. The other treatment effects have not been determined individually but their combined value has been obtained by subtracting the variances and covariances for (a) and (b) from the composite values for dose of vitamin D in Table XL. The reduced sum of squares for error is the same as that in Table XL.

The intermediate steps in computing the reduced treatment sums of squares are shown in the lower part of Table XLII. The entries in each of the first three rows of the table were added to the corresponding values in the row for error. The variation in each $[y^2]$ accounted for by the slope of y on v was then computed, that for Error + S vs. U being $B_v^2 = .02961^2/.06035 = .01453$. This term with 1 degree of freedom was then subtracted from $[y^2]$, giving $.16579 - .01453 = .15126$ with 25 degrees of freedom. To isolate the treatment effect, the reduced sum of squares for error was also subtracted, leaving $D^2 = .15126 - .03388 = .11738$, the required value. The other reduced sums of squares were obtained similarly. As was to be expected, the reduced values for both D^2 and B^2 were less than the approximations based on the adjusted T_a' and T_b'. Except for (a) and (b) the treatment effects were not significant.

The standard error of M was recomputed with the reduced values of D^2 and B^2, giving $s_M = .02332$ instead of .02318 as before. The confidence limits were computed by Eq. 34a with $C^2 = 1.0091$ to obtain $X_L = .2010 \pm .0483$ or 178 and 222 units per gram at $P = .05$. The confidence limits determined by Finney's method (38) were $X_L = .2013 \pm .0503$, corresponding to 177 and 223 units per gram.

VI. Assays Where the Variation in Response Is a Function of the Dose

In all except the slope-ratio assays the response for computing the vitamin content should meet two objectives. First, it should plot as a straight line against the log dose over the widest possible range and, secondly, it should be equally variable at all dosage levels. In most cases, including those which have been considered so far, both requirements can be met by a suitable measure of the response. In a few cases, how-

ever, the unit which leads to a linear dosage-response line is more variable at some dosage levels than at others. When such a line is computed, the unequal variation should be adjusted by the use of weighting coefficients which are inversely proportional to the variance expected at each dosage level. Observations with a larger variance and therefore a smaller reliability are given less weight than those where the variability is less. Several statistical methods have been proposed for handling this problem, which differ both in basic theory and in computational technique. In the analysis adopted here the weights have been derived by the method of maximum likelihood, a relatively advanced statistical technique, although the method of computation can be described in a simple form. Two types of vitamin assays fall in this category, those based upon the incidence of an all-or-none response and one which depends upon the reaction time to vitamin A. They will be considered separately.

1. The Analysis of Assays with an All-or-None Response

Many biological reactions are of an all-or-none type, such as measurements of toxicity from the percentage mortality. The test animals are assigned to several comparable lots, and all individuals in each lot are given the same dose of vitamin. The doses are selected so that most or all of them fall in the intermediate zone in which the percentage response of each lot is between 0 and 100. The percentage of animals which react is inherently less efficient than a graded response because the only information concerning each individual in a lot is whether or not it reacts. Nevertheless, in several vitamin assays an all-or-none reaction is both more convenient and more reliable. Vitamin E activity, for example, is determined from the percentage of positively mated female rats which produce a litter (p. 142). Other cases are the vitamin K assay from the incidence of normal blood clotting in chicks (p. 160), thiamine assays from the percentage cure of polyneuritis in rats (p. 197), in pigeons (p. 199) and in chicks (p. 200), and the vitamin A assay from the cure of xerophthalmia in rats (p. 87).

To determine the form of the dosage-response curve underlying these and similar assays, equivalent lots of test animals are given graded dosage levels of the vitamin supplement, so that the reaction in the successive lots will cover the range from 0 to 100%. A diagram of the percentage response against the log dose of vitamin then leads to a symmetrical sigmoid curve (Fig. 7, p. 147) which is asymptotic to 0 and 100%. As described on p. 146, most curves of this type may be identified with the cumulative normal or Gaussian curve. They are assumed to describe the variation in the susceptibility of individual animals to the

vitamin when susceptibility is measured in terms of the logarithm of the threshold dose. This assumption can be tested graphically by plotting the percentage response against the log dose on so-called probability paper, of which one variety is shown in Fig. 5. If the series of observations agrees with the hypothesis, the points plotted on probability paper should define a straight line.

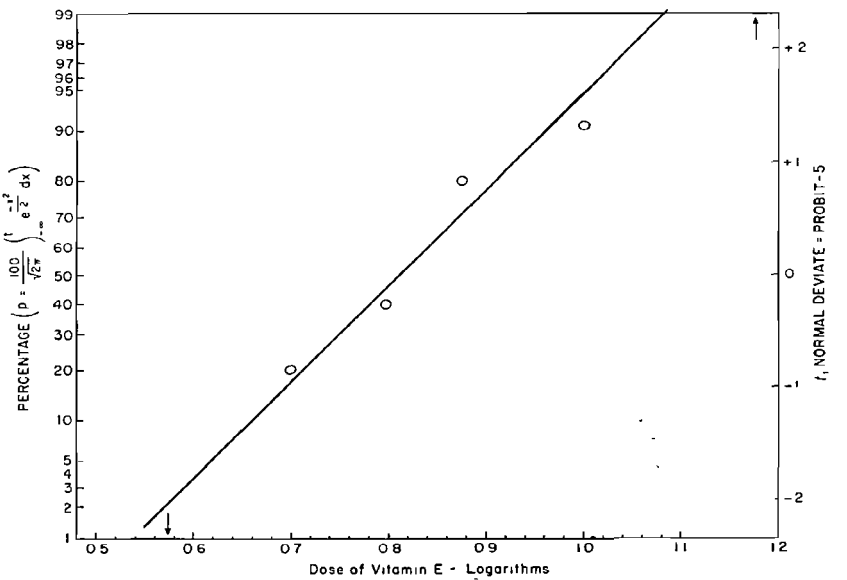

FIG. 5. Probability paper for plotting percentages directly on a probit scale, illustrated with the data in Table XLIII.

Instead of converting the graph paper to the probability scale, each percentage response can be transformed to the corresponding deviate of the normal curve in units of standard deviations. This has been termed the "normal equivalent deviation" or "N.E.D." by Gaddum (40). With the addition of 5 (to avoid minuses) it is known as the "probit," so that the probit for 50% response is 5. Table IX in Fisher and Yates (1) gives the probit corresponding to each percentage. The probits are plotted against the log dose on cross-section paper (ruled in millimeters or $\frac{1}{20}$ in.) and in the great majority of cases the plotted points can be fitted by a straight line. Since this diagram leads directly to the computed line, it is the one described here.

Two stages are involved in the standard method for evaluating all-

or-none experiments with probits (1, 41, 42, 43). A provisional line is fitted to each series of plotted points, either by inspection or by computation without weighting, using orthogonal coefficients when the log doses are spaced equally. This provisional estimate is required in the calculation of the weighted regression line in the second stage. The weighted regression line usually agrees so well with the provisional estimate that frequently the second stage is omitted. Hence the two stages will be considered separately.

A. THE PROVISIONAL DOSAGE-EFFECT CURVE

The first step in solving the provisional line is to convert doses to logarithms and each percentage response between 0 and 100% to tabular or empirical probits. The probits for observed percentages of 0 and 100% are indeterminate at this stage. In computing the curve they carry relatively little weight, so that they are often ignored in fitting the provisional line. Small vertical arrows below and above the line based on the intermediate points will indicate the direction of their influence on the computed curve. A straight line is fitted to the plotted points by inspection, using a transparent straightedge. If the doses have been spaced equally on a logarithmic scale, as is recommended, the slope of the line may be computed by the short-cut in Eq. 7 or 7a with T_d standing for a probit response and $f = 1$. The provisional line will pass through the unweighted mean probit and the middle log dose. It should be drawn on the diagram to check the arithmetic.

More often the provisional line is drawn by inspection. Its equation may be determined by interpolating two points from opposite ends X_1Y_1 and X_2Y_2. The slope of the straight line is estimated as

$$b = \frac{Y_1 - Y_2}{X_1 - X_2} \tag{51}$$

and its zero intercept as

$$a' = Y_1 - bX_1 \tag{52}$$

The expected probit Y for any given log dose X can be computed by Eq. 1a as $Y = a' + bX$, or, alternatively, the log dose expected at any given probit. Of the values interpolated from this line, one of the most frequent is the logarithm of the median effective dose or the log ED50. At this level $Y = 5$ and the equation may be rearranged as log ED50 $= (5 - a')/b$. The error in the estimate tends to be minimal at the log ED50 if the original observations are scattered more or less equally above and below 50%. The potency of the International Standard of vitamin E, for example, was defined in terms of the ED50.

An approximate standard error applying to the graphic estimate of

the log ED50 has been described by Litchfield and Fertig (44). It depends upon a more or less even scatter of the observations about 5 probits, with approximately the same number of individuals at each dosage level. The difference between the mean probit and 5 can then be neglected and the standard error computed from two terms, the standard deviation in units of the log dose and the sum of the weights used for the computed curve. Since the probit is measured in units of standard deviations, the reciprocal of the slope $1/b = \lambda$ is an estimate of the standard deviation of the individual effective log dose. The sum of the weights depends upon weighting coefficients which vary with the expected response, and the number of animals at each level. If we use only the observations at dosages corresponding to a range of 3.5 to 6.5 probits as interpolated from the provisional line, the average weighting coefficient (Z^2/PQ) is approximately one-half. The approximate standard error may then be computed as

$$\text{Approximate } s_{\log \text{ED50}} = \frac{\lambda}{\sqrt{N'/2}} = \frac{1}{b\sqrt{N'/2}} \tag{53}$$

where N' is the total number of individuals at all dosage levels in the restricted range. This approximation may be used with the provisional log ED50 for a preliminary comparison of different preparations.

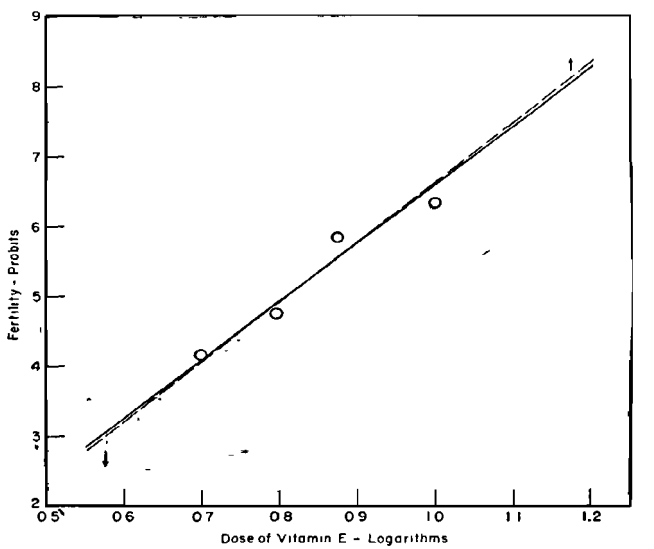

FIG. 6. Dosage-effect curve for vitamin E in Table XLIII, the broken line representing the graphic provisional curve and the solid line the computed curve.

TABLE XLIII. Estimation of the Dosage-Effect for a Concentrate of Vitamin E from the Percentage of Deliveries in Positive Matings

Data of K. E. Mason (45)

Data for provisional curve					Computation of weighted curve					
Single dose mg.	No. of rats N	Fertility rate %	Log dose x	Empirical probit	Expected probit Y	Weighting coefficient Z^2/PQ	Weight w	Working probit y	Intermediate products wx	wy
3.75	5	0	.574		2.98	.127	.6	2.56	.3444	1.536
5.0	10	20	.699	4.16	4.05	.455	4.6	4.16	3.2154	19.136
6.25	10	40	.796	4.75	4.87	.633	6.3	4.75	5.0148	29.925
7.5	10	80	.875	5.84	5.55	.570	5.7	5.82	4.9875	33.174
10.0	11	90.9	1.000	6.33	6.62	.231	2.5	6.26	2.5000	15.650
15.0	11	100	1.176		8.12	.010	.1	8.41	.1176	.841

Provisional line: $Y = -1.924 + 8.54X$

Computed dosage-effect curve:

$S(w) = 19.8$ $\quad [wx^2] = .21804$ (Eq. 58) $\quad [wy^2] = 16.096$ (Eq. 61)

$S(wx) = 16.1797$ $\quad [wxy] = 1.8145$ (Eq. 59) $\quad \chi^2 = .995, n = 2$ (Eq. 63)

$\bar{x} = .817157$ (Eq. 57) $\quad b = 8.3221$ (Eq. 60) $\quad t = 1.960$ (at $P = .05, n = \infty$)

$S(wy) = 100.262$ $\quad B^2 = 15.101$ (Eq. 62) $\quad B^2 - t^2 = 11.259$

$\bar{y} = 5.063737$ (Eq. 57) $\quad Y = -1.737 + 8.322X$

For $Y = 5$, $X = (5 + 1.737)/8.322 = .810$, and $s_x = \frac{1}{8.322}\sqrt{\frac{1}{19.8} + \frac{(.0637)^2}{11.259}} = .0271$ (Eq. 66)

Example 16a. A representative dosage-effect curve based upon an all-or-none response is that reported by Mason (45) on the relation between fertility rate and the dose of vitamin E concentrate. The fertility rate at six different dosage levels was measured in groups of 5 to 11 animals as shown in the first three columns of Table XLIII. In the next two columns the doses have been changed to logarithms and the four percentages in the intermediate zone to empirical or tabular probits with Table IX in reference (1). The probits were then plotted against the log dose in Fig. 6, indicating the responses of 0 and 100% with arrows. The broken line was placed by inspection and Y interpolated as 3.20 and 7.47 at $X = .6$ and 1.1 respectively. The equation of this provisional line was determined by Eqs. 51 and 52 as $Y = -1.924 + 8.54X$, from which

$$\log \text{ED50} = \frac{5 - (-1.924)}{8.54} = 0.811$$

From inspection of the broken line in Fig. 6, it is evident that the fertility in only three of the six sets had an "expected probit" within the range of 3.5 to 6.5 probits. These represented a total of $N' = 30$ rats within the range for computing the approximate standard error. Substituting in Eq. 53, the approximate $s_{\log \text{ED50}} = 1/8.54\sqrt{15} = .0302$. Hence the log ED50 and its error were estimated approximately as $.811 \pm .030$. In original units this was equivalent to $6.47 \pm .45$ mg. (Eq. 33) of vitamin E concentrate.

B. THE GRAPHIC ESTIMATION OF POTENCY

All-or-none assays may be evaluated by a simple extension of the graphic techniques for the dosage-effect curve. Each computed assay is essentially a comparison of two regression equations which have been determined with weights. The factorial designs which are so effective in simplifying most assays based upon a graded response cannot be used with precision in all-or-none assays. This is because the weight of each probit in an all-or-none assay depends not only upon the number of animals used in its determination but also upon the weighting coefficient for the response expected at the given dose. The expected response, in turn, may vary considerably from one experiment to the next at the same dosage level, so that the weighting coefficients used to adjust differences in inherent reliability cannot be predicted. In consequence, it is nearly impossible to allot in advance larger numbers of animals to the doses with the smaller weighting coefficients in order to equalize the weight of each probit.

Even though factorial coefficients cannot be used in its evaluation,

an all-or-none assay follows much the same pattern as a graded response assay. A provisional potency is assigned to the unknown which is then administered at the same dosage levels as the standard. Dosages should form a geometric series and, so far as is predictable, fall within the range giving a response between 10 and 90%. Because of the unpredictable shifts in sensitivity and the vital importance of having at least two responses in the intermediate zone, three or more dosage levels are often preferred.

Just as with the dosage-effect curve, the calculation of an all-or-none assay involves a provisional estimate, usually graphic, followed by a computed log potency and its error. The graphic stage has now been provided with nomographs and other aids necessary for a complete solution (46, 47) although these are still limited by the lesser precision of a provisional line. Here the full analysis will be reserved for the computed estimate, which with a calculator is not onerous, and the graphic estimate limited to simpler terms.

The simple graphic method, which is discussed for vitamin E on pp. 148-151, may be summarized. Each response in the assay between 0 and 100% is converted to probits and plotted on decimal cross-section paper against the corresponding log dose of standard and unknown. Observations at 0 and 100% are indicated by vertical arrows below and above the trend. With the aid of a transparent triangle and a straightedge along which it may be moved, two parallel lines are fitted by inspection to the plotted points, one for the standard and one for the unknown. The horizontal distance between these two lines, as interpolated from the diagram, is an estimate of the log relative potency (M) of the unknown. This is done most readily by interpolating the log dose from each curve (X_s and X_u) at some selected level of response, such as probit 5, and computing

$$M = X_s - X_u \tag{23c}$$

The standard error of this log potency can be approximated by an extension of Eq. 53 giving

$$s_M = \frac{1}{b}\sqrt{\frac{2}{N'_s} + \frac{2}{N'_u}} \tag{54}$$

where N'_s and N'_u are the total number of animals on doses of the standard and of the unknown, respectively, which have an expected response between 3.5 and 6.5 probits inclusive, as read from the fitted lines.

Example 17a. The graphic technique is illustrated by the antisterility assay in Table LXVII (p. 615). The results in the last two columns have been plotted in Fig. 12 (p. 615) and fitted by inspection with two parallel lines. The log dose at probit 5 was interpolated from the two lines as

$x_s = 0.75$ and $x_u = 0.94$, from which $M = 0.75 - 0.94 = -0.19$ by Eq. 23c. In computing s_M, the only dose for which the interpolated response fell outside the range from 3.5 to 6.5 probits was the low dose of the unknown, which gave a response of 0%. With this omission, $N'_s = 24$ and $N'_u = 22$. The dosage-effect curve for the standard was used in computing the slope by Eq. 51 to obtain $b = (6.45 - 3.55)/(1.0 - 0.5) = 5.8$. By substitution in Eq. 54,

$$s_M = \frac{1}{5.8}\sqrt{\frac{2}{24} + \frac{2}{22}} = .072$$

The log potency of the unknown has been estimated as $M = -0.19 \pm .072$, from which, by Eq. 33, the potency of the unknown was equal approximately to 65 ± 11% that of the standard.

C. THE COMPUTED DOSAGE-EFFECT CURVE

There are a number of advantages in computing the dosage-effect curve. The provisional equation is replaced by one which makes the most of the original observations and permits a test of the agreement between the observed responses and the hypothesis used in the probit transformation. Series in which the observations are not scattered evenly about 50% can be handled effectively. Of especial importance, the data from several similar small experiments often can be combined to obtain a sufficiently good estimate of the slope to lead to a positive result. The computation is one of successive approximations, starting with the graphic provisional line in the preceding section. If the first computed estimate differs from it but little, as is usually the case, the computed line is accepted as final. The dosage-effect curve is fitted by least squares but with two modifications. Each observation is assigned a weight and the empirical probits are replaced by "working probits."

The weights are based upon the variance of the probit response. The probit is here a transformed percentage response and each original percentage ($100p$) can be considered a sample from a binomial distribution. In consequence its variance is a function of the response expected at that dose, so that $V(p) = PQ/N$, where P is the expected value of the observed p and $Q = 1 - P$. The variance of P is a maximum at $P = Q = 0.5$ and decreases toward 0 and 1.00. In transforming percentages to probits, this relation is reversed so that the variance is a minimum at the log dose corresponding to 5 probits on the curve and increases progressively above and below this value. To correct this change each probit is given a weight (w) equal to the reciprocal of its variance or

$$w = N\frac{Z^2}{PQ} \tag{55}$$

where Z is the ordinate of the normal curve at the expected proportionate response P. The fractional term is known as the weighting coefficient and has been tabled for different expected probits (1, Table XI). It is .637 for 5 probits or 50%, .439 for 84.1% or 6 probits, and .131 for 97.7% or 7 probits. The weights at values less than 50% parallel those for more than 50% and hence are the same for 4 (15.9%) and 6 and for 3 (2.3%) and 7 probits.

The empirical probits based upon large sample theory have the limitation that no value is available for an observed 0 or 100% response. These carry reduced weight because of their small weighting coefficients, but in some cases may contribute appreciably to the estimated dosage-response curve. Moreover, at percentages below 10% or above 90% or with small groups, the empirical probits based upon the theoretical curve may prove somewhat erratic. This limitation is met by substituting working probits for empirical probits, especially at the ends of the curve. The working (or corrected) probits can be computed readily for all observations with the tables now available (such as 1, Table XI). In addition to the table one needs the actual percentage response and an estimate of the expected probit (Y) from the provisional line. Each working probit is computed from the equation

$$y = \left(Y - \frac{P}{Z}\right) + p\left(\frac{1}{Z}\right) = \left(Y + \frac{Q}{Z}\right) - q\left(\frac{1}{Z}\right) \qquad (56)$$

where p is the proportion which was observed to react and P and Z are the proportionate area and the ordinate of the normal curve corresponding to the provisional expected probit Y. Some tables (1, Table XI) [3] give the value for both the maximum working probit ($Y + Q/Z$) and the minimum working probit ($Y - P/Z$) where $P + Q = p + q = 1$. These terms and the range $(\frac{1}{Z})$ are tabled for different values of Y, so that it is a simple matter to compute the working probit, y, for each observation. In the central part of the curve or in the case of observations which differ but little from the line, the working probit usually agrees very closely with the empirical value. The working probit, however, is always determinate, even when the observed p is 0 or 1.

The corrected dosage-effect curve is then computed from the values for w, x, and y. The work is facilitated by computing and recording the products wx and wy for each observation. The remaining calcula-

[3] A more detailed table which makes interpolation unnecessary has been published by D. J. Finney and and W. L. Stevens in *Biometrika* (35, 191, 1948) and can be purchased separately.

tion resembles that described earlier in this chapter. The weighted means are determined to 6 or 7 decimal places as

$$\bar{x} = \frac{S(wx)}{S(w)} \quad \text{and} \quad \bar{y} = \frac{S(wy)}{S(w)} \tag{57}$$

Cross-multiplying columns and accumulating the products in the calculator, the weighted sums of squares and products of the deviations from the two means are determined as

$$[wx^2] = S(wx^2) - \bar{x}S(wx) \tag{58}$$

and

$$[wxy] = S(wxy) - \bar{y}S(wx) \tag{59}$$

The slope of the dosage-effect curve is then equal to

$$b = \frac{[wxy]}{[wx^2]} \tag{60}$$

If the computed equation (1a) agrees reasonably well with its provisional estimate, it is accepted as final. Should it disagree materially, however, it also is considered as provisional, and the calculation is carried to a second approximation.

Agreement of the observations with the computed line can be tested by χ^2, one of the most important of the statistical tests of significance. χ^2 may be computed as the ratio of an observed to an expected variance. Since each observation is weighted by the reciprocal of its expected variance, the sum of the squared deviations about the fitted line follows the χ^2 distribution. The weighted total variation in the response is computed as

$$[wy^2] = S(wy^2) - \bar{y}S(wy) \tag{61}$$

By analogy with Eq. 9 the effect of the straight line is estimated as

$$B^2 = \frac{[wxy]^2}{[wx^2]} \tag{62}$$

leading to

$$\chi^2 = [wy^2] - B^2 \tag{63}$$

The values of χ^2 for various probabilities of random occurrence depend upon the degrees of freedom in each and are given in Table IV of reference (1). To test whether the observations agree with a specific curve, χ^2 is entered into this or a similar table with $k - 2$ degrees of freedom, where k is the number of groups in which the expected number in the smaller class, either positive or negative, is equal to 0.5 individual or more. When the observations scatter about the dosage-effect curve

no more widely than would be expected by chance at $P = .05$, the data are considered consistent with the fitted line.

When there is such agreement, the errors about the curve may be generalized. The observed variation χ^2/n ($\approx s^2$) is replaced by that expected in a large series of similar tests or by $n/n = 1$. The standard errors of position and of slope are then computed as

$$s_a = \sqrt{1/S(w)} \tag{64}$$

and

$$s_b = \sqrt{1/[wx^2]} \tag{65}$$

which may be used in the same way as those from Eqs. 12 and 13 for estimating the variation about the dosage-effect line. In computing the limits corresponding to a given probability, all standard errors for a dosage-effect curve from homogenous data are used with a deviate of the normal distribution or the value of t for $n = \infty$ so that at $P = .05$, $t = 1.96$. The standard error of a log dose X computed from the line for any given Y such as $Y = 5$, is determined by

$$s_X = \lambda \sqrt{\frac{1}{S(w)} + \frac{(Y - \bar{y})^2}{B^2 - t^2}} \tag{66}$$

where $\lambda = 1/b$. The confidence limits for X at a given Y are

$$X_L = \bar{x} + \lambda C^2 (Y - \bar{y}) \pm tCs_X \tag{67}$$

where $C^2 = B^2/(B^2 - t^2)$.

Example 16b. The calculation of the vitamin E experiment in Table XLIII may be extended from the provisional curve to the first computed estimate as shown in the last part of the table. The provisional equation was used with each observed value of x to obtain the expected probits Y in the sixth column. The weighting coefficients corresponding to these expected probits are given in the next column and each was multiplied by N to obtain the weights w (Eq. 55). Although the individual groups differed in size by little more than twofold, the weights varied more than sixty-fold. The working probits y were computed by Eq. 56 from the observed percentages and the expected probits Y for all values in the series. Three of the four working probits in the essential intermediate zone agreed well with the empirical probits which they replaced. The last two columns of intermediate products were required for computing the slope and χ^2.

Substitution in Eqs. 57 to 63 gave the results shown beneath the table. From the percentages corresponding to the expected probits, less than 0.5 individual was expected at the smallest and largest dosages.

Hence the estimated $\chi^2 = 0.995$ would be assigned 2 degrees of freedom. It showed that the observations agreed with the computed curve well within the sampling error. The computed slope of $b = 8.32 \pm 2.41$ differed so little from its provisional value (Fig. 6) that the first computed approximation sufficed. The log ED50 and its standard error (Eq. 66) were 0.810 ± 0.027. In original units the median effective dose of 6.45 ± 0.40 mg. represented little change from the previous estimate of 6.47 ± 0.45 based on the provisional curve.

D. THE COMPUTED POTENCY AND ITS PRECISION

All-or-none assays are similar to assays from two dosage-response curves based upon a graded reaction as described on pp. 487-489. As in the other case, computation tests whether the use of a combined assay slope is justified and provides a more objective measure of the assayed potency and its precision. The procedure is one of computing two dosage-effect curves by the same methods as in the preceding section, except that the provisional lines used in estimating weights and working probits are made parallel.

With this one exception, separate curves are computed as described for the standard and for the unknown. The variability about each line is tested by χ^2. If the sum of the two χ^2's is not significantly greater than would be expected for the total of their respective degrees of freedom, the observations can be assumed to represent straight lines within the sampling error in terms of probits and log dose. The combined or assay slope is then determined from the numerators and denominators of the composite slopes as

$$b_c = \frac{[wxy]_s + [wxy]_u}{[wx^2]_s + [wx^2]_u} = \frac{S[wxy]}{S[wx^2]} \tag{24a}$$

The total variation in y accounted for by b_c is

$$B_c^2 = \frac{S^2[wxy]}{S[wx^2]} \tag{25a}$$

The discrepancy in slope between the separate component curves in the assay is tested by

$$\chi_b^2 = B_s^2 + B_u^2 - B_c^2 = S(B_i^2) - B_c^2 \tag{68}$$

where the subscript i refers successively to the standard (s) and to the one or more unknowns (u). χ_b^2 has one degree of freedom less than the number of slopes involved in the comparison. For a valid assay it should not exceed the tabular value of χ^2 (1, Table IV) at

$P = .05$. It is sometimes convenient to combine the several components of χ^2 to obtain

$$\text{Assay } \chi^2 = [wy^2]_s + [wy^2]_u - B_c^2 = S[wy^2]_i - B_c^2 \qquad (69)$$

where each $[wy]^2$ is computed by Eq. 61 and the assay χ^2 has one less degree of freedom than in $S[wy^2]_i$.

The log ratio of potencies is computed from the means for the log dose and probit response by Eq. 23a, just as in an assay with a graded response. When the first computed values of b_c and of M differ but little from their provisional estimates, they are generally taken as final. In cases where the difference is substantial, however, the first computed values are used as provisional values to obtain new weights and working probits from which a second solution is calculated. The process may be continued to a third or fourth estimate. In most cases, however, successive approximations converge rapidly, so that more than two computed estimates are rarely needed.

If χ^2 for the assay has shown agreement within the sampling error, the standard error of M can be computed with $s^2 = 1$ as noted in the preceding section to obtain

$$s_M = \frac{1}{b_c}\sqrt{\frac{1}{S(w)_s} + \frac{1}{S(w)_u} + \frac{(\bar{y}_s - \bar{y}_u)^2}{B_c^2 - t^2}} \qquad (70)$$

where t is the value for $n = \infty$ or the deviate of the normal curve. Confidence or fiducial limits, in turn, may be determined by Eq. 34, in which case C is computed from $C^2 = B_c^2/(B_c^2 - t^2)$.

The assay χ^2 (Eq. 69) for an occasional experiment may exceed its expected value significantly. If an examination of the plotted data shows excessive scatter but no clear-cut departure from linearity, s_M and the confidence limits may be computed by Eqs. 32 and 34, with $N_u = S(w)_u$, $N_s = S(w)_s$, $s^2 =$ assay χ^2/n and t for the degrees of freedom in the assay χ^2.

Example 17b. The agreement of a calculated estimate of potency with that determined graphically may be shown conveniently with the all-or-none assay of synthetic DL, α-tocopherol (the unknown) in terms of its acetate (the standard). The graphic analysis in Fig. 12 from the data in Table LXVII has been provided with an estimate of its error on p. 542. The values used for the graphic estimate have been repeated in the first six columns in the upper part of Table XLIV. The log doses for the standard and for the unknown were interpolated from Fig. 9 at $Y = 5$ as $X_s = 0.75$ and $X_u = 0.94$, from which $a'_s = 5.00 - 5.8 \times 0.75 = 0.650$ and $a'_u = -0.452$ by Eq. 52. With these terms and

TABLE XLIV. Calculation of Potency in Antisterility Assay for Vitamin E Potency from Data in Table LXVII, first computed estimate

Vitamin	Dose mg.	No. of rats N	Fertile per cent	1 + log dose, x	Empirical probit	Expected probit Y	Weight w	Working probit y	Intermediate products wx	Intermediate products wy
S	0.4	8	25	.602	4.33	4.14	3.9	4.34	2.3478	16.926
	0.6	8	50	.778	5.00	5.16	5.0	5.00	3.8900	25.000
	0.9	8	100	.954		6.18	3.0	6.78	2.8620	20.340
				Total			11.9		9.0998	62.266
U	0.4	10	0	.602		3.04	1.4	2.61	.8428	3.654
	0.6	12	16.7	.778	4.03	4.06	5.5	4.03	4.2790	22.165
	0.9	10	60	.954	5.25	5.08	6.4	5.25	6.1056	33.600
				Total			13.3		11.2274	59.419

Expected probits: $Y_s = .65 + 5.8x$; $Y_u = -.452 + 5.8x$.

Term	Standard	Unknown	Both
$\bar{x}$	.7646891	.8441654	
$\bar{y}$	5.2324370	4.4675940	
$[wx^2]$	.21163	.18339	.39502
$[wxy]$	1.42968	1.33901	2.76869
b	6.7556	7.3014	7.0090[a]
B^2	9.6583	9.7767	19.4057[a]
$[wy^2]$	10.5611	9.8019	20.3630
χ^2	.9028	.0252	.9573[a]

$\chi_b^2 = .0293$ (Eq. 68), $\bar{y}_u - \bar{y}_s = -.76485$

$M = -.1886$ (Eq. 23a), $t = 1.960$ ($P = .05$, $n = \infty$)

$B^2 - t^2 = 15.5641$, $s_M = .0633$ (Eq. 70)

[a] Computed from other entries in same column. other values are sums of the first two entries in the same row.

$b = 5.8$, the expected probits in the seventh column were computed for each log dose of both standard and unknown. The expected probits were needed in computing the weights w and the working probits y in the next two columns with the aid of Table XI in reference (1). In this table the expected probits are given only to the nearest 0.1, so that linear interpolation is required. With a calculator this is an easy operation. The weighting coefficient for $Y = 4.14$, for example, is $.6(.471) + .4(.503) = .484$, which is multiplied by $N = 8$ to obtain $w = 8 \times .484 = 3.9$. Similarly, the weighting coefficient for $Y = 5.16$ is $.4(.634) + .6(.627) = .630$, and $w = 8 \times .630 = 5.0$. The others were obtained similarly. In the present case the empirical or tabular probits for the percentages between 0 and 100 differed little from their "expected" values. The greatest discrepancy was in the first row of the table, where for $Y = 4.14$ the minimum working probit ($Y - P/Z$) was .6(3.408)

+ .4(3.469) = 3.432, the range ($1/Z$) was .6(3.758) + .4(3.452) = 3.636, and the working probit by Eq. 56 was y = 3.432 + .25 × 3.636 = 4.34. The intermediate products wx and wy completed the first step of the calculation as shown in the upper part of Table XLIV.

The next step was to compute the dosage-effect curves separately for the standard and the unknown as shown in the lower part of Table XLIV following the procedure described in Table XLIII. Both χ^2's in the last row, each with not more than one degree of freedom, indicated good agreement with the fitted lines. The two lines were substantially parallel as judged by χ_b^2 = 9.658 + 9.777 − 19.406 = .029 with one degree of freedom, so that the log potency has been computed as M = .7647 − .8442 + (4.4676 − 5.2324)/7.009 = −0.189. Its standard error by Eq. 70 was s_M = 0.063. These values may be compared with the graphic estimate from the same original data of M = −0.19 ± 0.072.

Because of the small difference between the graphic and computed potencies, the calculation ordinarily would be terminated at this stage. However, the computed assay slope (b_c = 7.009) was over 20% greater than its provisional value. To illustrate the convergence in successive approximations when one-third of the doses gave 0 or 100% effect, the calculation has been carried to a second and third computed estimate with the following results:

Estimate	Assay slope	M	s_M	Assay χ^2
Graphic	5.80	—0.19	.072	
1st computed	7.009	—0.1886	.0633	.957
2nd computed	7.345	—0.1904	.0617	1.414
3rd computed	7.364	—0.1914	.0619	1.560

Convergence is usually retarded by observed responses of 0 and 100%, but even with this handicap the first computed estimate was adequate. The assay χ^2 tended to increase in successive approximations, but not seriously. It had two degrees of freedom if we do not count the observed 0% which contributed progressively less to the computation. Since in statistical jargon the curves have been computed so as to maximize the likelihood rather than to minimize χ^2, this was not unexpected.

2. A Graded Response with Unequal Variance

An assay for vitamin A has been proposed (48) in which the complete analysis resembles the maximum likelihood solution for an all-or-none response. Each dose of vitamin A is administered over two days in four equal fractions to depleted ovariectomized rats. The response

is taken as the number of days required from the start of dosing for the cellular contents of the vagina to change from the squamous cells characteristic of the depleted rat to leucocytes or to a mixture of leucocytes, epithelial and squamous cells, and return to the original depleted condition. In days the reaction time is equally variable at all dosage levels but when plotted against the log dose of vitamin, leads to a concave curve. The logarithm of the reaction time, on the other hand, gives a straight line over a tenfold range of doses, but its variability is larger at the lower than at the upper dosage levels. For an exact solution the logarithm of the average reaction time is adopted as the criterion of response, and each observation is assigned a weight inversely proportional to its expected variance. For many assays, however, the potency and its error can be approximated sufficiently well by a shortened method which weights each observation equally.

A. THE DOSAGE-RESPONSE CURVE

The method of analysis differs from the procedure used with probits in that the observed variance in original units is an integral part of each weight. The first stage is to compute in days the mean response $\bar{y}'$ for each dosage level and the sum of squares of the deviations about each mean $[y'^2]$. The sums of squares within groups are totaled $S[y'^2]$ over all dosage levels and divided by the total degrees of freedom n_c to obtain a combined estimate of the variance (s_c^2) within groups. The homogeneity of the components entering into the pooled variance can be tested by χ^2 (49). χ^2 is computed from the logarithms of the variances at each dosage level by the equation

$$\chi^2 = \frac{2.303}{C'} \left\{ n_c \log s_c^2 - S(n \log s^2) \right\} \tag{71}$$

where s_c^2 is based upon the total sum of squares over all groups with n_c degrees of freedom, s^2 is the variance of each group with n degrees of freedom and C' is defined as

$$C' = 1 + \frac{1}{3(k-1)} \left\{ S\left(\frac{1}{n}\right) - \frac{1}{n_c} \right\} \tag{72}$$

Given k variances, χ^2 has $k - 1$ degrees of freedom. The significance of χ^2 is determined by reference to the distribution of χ^2 such as that in Table IV, reference (1). When the variances observed at different doses of vitamin A with this experimental technique have been computed, they have proved to be homogeneous when the reaction time was measured in days.

Each mean response in days ($\bar{y}'$) is then converted to its logarithm

and log $\bar{y}'$ is plotted against the corresponding log dose of vitamin. A provisional line is drawn through these points by inspection and has the same purpose as in the preceding section. Its equation is determined from two interpolated points by Eqs. 51 and 52, and the expected Y is computed for each observed log dose. The weight of each observation is given by the equation

$$w = \frac{f(2.303m)^2}{s_c^2} \tag{73}$$

where m is the antilogarithm of Y, f is the frequency or number of observations at a given dose and s_c^2 is the combined estimate of the variance within groups in original units (days). It is useful to compute the curve with an abbreviated working weight w' which includes only those terms in Eq. 73 that vary from dose to dose. If the number of observations (f) is constant at each dosage level, we may use m^2 as a working weight and correct for the missing components at the end of the calculation.

The logarithm of the mean response in days is equivalent in a sense to an empirical probit. If the data were quite irregular it would not lead to the maximum likelihood estimate of the computed line. By a procedure analogous to that for working probits, a working response is determined as

$$y = Y + \frac{(\bar{y}' - m)}{2.303m} \tag{74}$$

The dosage-response curve can be computed from the values for x, y, and w by the Eqs. 57 to 60 in the preceding section. If the dosage interval is constant on a logarithmic scale, it is convenient to code the doses with the integers 0, 1, 2, 3, · · · and correct for the code at the end of the computation. The slope (b) in coded units is divided by the dosage interval (i) in logarithmic units to convert the slope to terms of y and the log dose.

Agreement of the observations with the fitted line can be tested by χ^2 after correction for the terms omitted from the weights (Eq. 73) in computing the curve. Using the working weights, we first determine the weighted sum of squares of the deviations in y, $[w'y^2]$ by Eq. 61 in the preceding section. The effect of slope (B^2) is computed similarly from Eq. 62. If the number of responses (f) has been the same at each dosage level, so that f could be omitted from the working weight, the agreement of y with the computed line is tested by

$$\chi^2 = \frac{([w'y^2] - B^2)\ 2.303^2 f}{s_c^2} \tag{75}$$

χ^2 has $k - 2$ degrees of freedom where k is the number of dosage levels.

The standard errors of the dosage-response curve depend upon whether the observations agree with the fitted straight line. Assuming such agreement, the standard error for position s_a is computed by Eq. 64, where $S(w)$ is the sum of the true weights rather than of the working weights. The variance in slope is computed similarly as the reciprocal of $[wx^2]$ by Eq. 65. If the dosage-response curve has been computed with working weights (w') and with coded log doses (x_1),

$$[wx^2] = \frac{2.303^2 fi^2 [w'x_1^2]}{s_c^2} \tag{76}$$

which applies in case the frequency is the same at all dosage levels.

The observed values of log $\bar{y}'$ sometimes agree so well with the graphic provisional line that the calculation may be based entirely upon the observed mean responses with little loss in precision. In this case the weight of each observation in Eq. 73 is determined with $\bar{y}'$ instead of m and each working response y, as defined in Eq. 74, is replaced by log $\bar{y}'$. The dosage-response curve is then computed without other change, a single calculation leading directly to the final line. When the observations agree so closely with the provisional line that this form of calcula-

TABLE XLV. Dosage-Response Curve for the Reaction Time of Rats to Vitamin A as Determined by Vaginal Smears

Data from Pugsley *et al.* (48)

	Frequency (f) of reaction time in days (y') at given dose of vitamin A in I.U.											
	50.0		80.0		128.0		204.8		327.7		524.3	
	y'	f	y'	f	y'	f	y'	f	y'	f	y'	f
	6	1	9	2	12	1	16	2	23	1	28	2
Frequency	7	1	10	1	13		17	1	24	7	29	1
Distribution	8	3	11	2	14	5	18	2	25	5	30	4
	9	3	12	5	15	2	19	2	26		31	3
	10	2	13	4	16	4	20	5	27	1	32	3
	11	4			17	2	21	2			33	1
$S(fy')$	128		162		210		265		343		427	
$\bar{y}'$	9.14		11.57		15.00		18.93		24.50		30.50	
$[y'^2]$	33.71		25.43		26.00		36.93		11.50		29.50	
s^2	2.593		1.956		2.000		2.841		.885		2.269	
log s^2	.4138		.2914		.3010		.4535		−.0531		.3558	

$S[y'^2] = 163.07$, $n = 78$, $s_c^2 = 2.0906$, log $s_c^2 = .32027$

$$\chi^2 = \frac{2.303}{1.0299} \{78 \times .32027 - 13 \times 1.7624\} = 4.629 \quad \text{(Eq. 71)}, \; n = 5$$

tion is justified, χ^2 as defined by Eq. 75 should not fall in the zone where the abbreviated calculation could affect its value critically.

Example 18. The computation may be illustrated by the data for the dosage-response curve from the original paper of Pugsley *et al.* (48). Six dosage levels of vitamin A varying from 50 to 524 International Units (I.U.) were prepared from the Canadian Standard Reference Oil by dilution with coconut oil. Each dose was 1.6 times that of the next smaller dose. The responses of 14 rats treated at each dosage level are shown by the frequency distributions in Table XLV. In the lower part of the table the mean $\bar{y}'$, the sum of squares $[y'^2]$, and the variance s^2 have been computed for each dosage level. There was no obvious relation between the variance within groups and the dose of vitamin A. The homogeneity of the variances was confirmed by the χ^2 test beneath the table ($P = .46$). In consequence $s_c^2 = 2.0906$ with 78 degrees of freedom could be computed from the pooled values of $[y'^2]$.

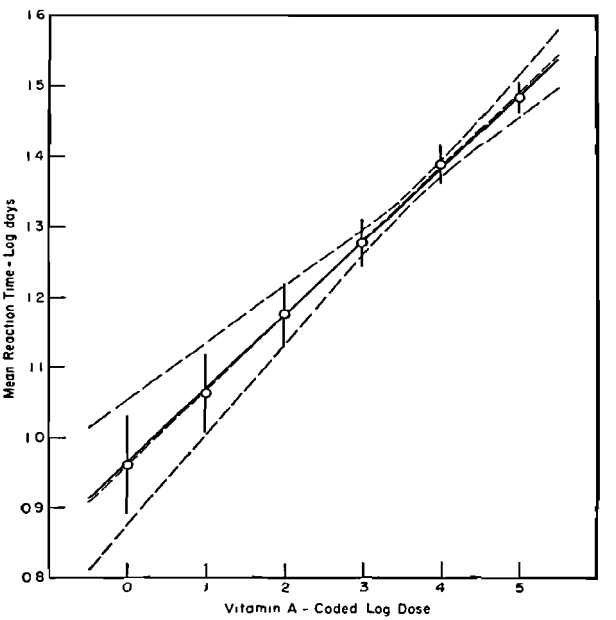

FIG. 7. Dosage-response curve for effect of vitamin A upon reaction time of ovariectomized rats from data in Table XLV. The broken straight line is the provisional curve and the solid line the computed curve, which has confidence limits at $P = .05$ shown by the two curved lines. Vertical bars represent one expected standard deviation for one rat in log days above and below the logarithm of the observed mean reaction time in days.

The calculation of the dosage-response curve is shown in Table XLVI. Since the interval between successive log doses was constant, they have been coded as x_1. The logarithm of each mean response (y) in the fourth column has been plotted against the coded doses in Fig. 7, and fitted by inspection with a straight line. It is apparent that the observations agreed well with the provisional curve. Accordingly, the calculation has been abbreviated by substituting y' for m in computing the weights with Eq. 73. Since the number of observations was constant at each dosage level, a working weight of $w' = \bar{y}'^2$ was used. Note that the weight assigned the response at the largest dose was more than ten times as large as that at the smallest dose. The weights adjusted the differences in the standard deviation of an individual log reaction time at each dose, plotted as vertical lines in Fig. 7. Corresponding to the approximation in the weights, the response was taken as $y = \log \bar{y}'$ for computing the curve. The intermediate products of the working weights with the coded doses and with the responses completed the table.

TABLE XLVI. Calculation of the Dosage-Response Curve from the Data in Table XLV

Dose I.U. per rat	Mean response in-days $\bar{y}'$	Coded log dose x_1	log y' y	Working weight $(= \bar{y}'^2)$ w'	Intermediate products $w'x_1$	$w'y$
50	9.14	0	.961	84	0	80.724
80	11.57	1	1.063	134	134	142.442
128	15.00	2	1.176	225	450	264.600
204.8	18.93	3	1.277	358	1074	457.166
327.7	24.50	4	1.389	600	2400	833.400
524.3	30.50	5	1.484	930	4650	1380.120
Total				2331	8708	3158.452
Mean					3.735736	1.354977

Provisional graphic line:

$$\text{at } x_1 = 0,\ y = .960, \qquad b = \frac{1.490 - .960}{5 - 0} = .106$$

$$\text{at } x_1 = 5,\ y = 1.490, \qquad a' = .960$$

Computed with coded log doses (x_1) and working weights (w'):

$[w'x_1^2] = 4575.21$ (Eq. 58) $\qquad B^2 = 49.9809$ (Eq. 62)

$[w'x_1y] = 478.198$ (Eq. 59) $\qquad [w'y^2] = 50.0221$ (Eq. 61)

$b = .10452$ (Eq. 60) $\qquad [w'y^2] - B^2 = .0412$

$a' = .9645$

In terms of log dose x:

$i = .204$ $\qquad \bar{x} = \log(50) + .204(3.7357) = 2.4611$

$b = .1045/.204 = .5123$ $\qquad Y = 1.3550 + .5123(X - 2.4611)$

The calculation in coded log doses with the working weights by Eqs. 58 to 62 is shown beneath the table. The log-dose interval of $i = 0.204$ was divided into the slope in code to obtain $b = .5123$ in units of log dose. After adjusting the mean dose to the same terms, the equation of the line after decoding was $Y = 1.3550 + .5123\ (X - 2.4611)$. When computed by the full maximum likelihood method $Y = 1.3562 + .5118\ (X - 2.4638)$, in substantial agreement with the result obtained in Table XLVI.

The agreement of the observations with the fitted line has been checked by χ^2. The working weights used in computing the dosage-response curve were too small by the factor $2.303^2 f/s_c^2 = 35.52$. Substituting this term in Eq. 75, $\chi^2 = .0412 \times 35.52 = 1.463$ with four degrees of freedom, for which P is about 0.83. To obtain the standard errors in position and slope, the sum of the working weights was multiplied by the same factor to obtain $S(w) = 35.52 \times 2331 = 82797$ and $s_a = \sqrt{1/82797} = .003475$. The standard error of the slope was computed similarly from Eqs. 76 and 65 as $s_b = .01216$. These standard errors have been used in Eq. 14 to compute the confidence limits at $P = .05$ for the dosage-response curve for vitamin A (Fig. 7).

B. THE ESTIMATION OF POTENCY

When the variance is related to the dose as in the present case, the vitamin potency may be computed by the method of maximum likelihood as in the all-or-none assay. An approximate estimation of potency is also available which is further simplified from the abbreviated calculation described above for the dosage-response curve. As compared with an all-or-none assay, it has the advantage of being uncomplicated by 0 and 100% responses and is therefore suitable for the factorial analysis described for other graded response assays. The first step in the calculation is the same for both the shortened and the exact analyses. Because of its greater efficiency, only the factorial design will be considered.

The reaction time to vitamin A in days (y') is recorded for each rat, and the first step is to compute the mean response ($\bar{y}'$) at each dose of vitamin. Each mean response in days is then converted to its logarithm, $y = \log \bar{y}'$, and this is the basic unit in the shortened solution. In addition, the sum of squared deviations [y'^2] at each dosage level is computed from the reaction time in days. In case their homogeneity is in doubt, the variances from these sums may be tested by χ^2 as defined in Eq. 71. All the sums of squares within doses are then added and divided by the total degrees of freedom to find s_c^2, the variance in days within doses. This term may be used in either the shortened or the exact solution. Alternatively, the calculation of the sums of squares and of the variance

can be avoided in the short method by estimating the standard deviation from the mean range. The range is determined at each dosage level and averaged to obtain $\overline{R}$, which, in turn, is multiplied by twice the coefficient for $\overline{R}_y$ in Table XXXII to obtain s_c.

The shortened computation of log potency and its error gives equal weight to the response at each dosage level. M' is determined by Eqs. 29 or 30 for two-dose or three-dose factorial assays respectively, computing T_a and T_b in the usual way from the 4 or 6 values of y. Its standard error s_M is determined approximately by Eq. 32a from which the term $D^2/(B^2 - s^2t^2)$ is omitted. Since $f = 1$, the equation for the standard error can be reduced to

$$s_M = \frac{Kis_c}{T_b\sqrt{f}\,(\text{antilog}\,\overline{y})} \tag{77}$$

where $K = 0.8686$ for a two-dose assay and 1.418 for a three-dose assay, $\overline{y}$ is the average of the 4 or 6 values of y, and the other terms have their previous significance. For the standard deviation in days within doses s_c, either the efficient estimate based upon the total sum of squares within doses may be used or the approximate value based upon the mean range.

The exact solution requires unequal weighting of the responses at each dosage level. If the experiment is planned factorially, with the same number of observations at each dosage level and the same interval in logarithms between successive doses, it is preferable to carry out the computation with coded doses x_1 and with working weights, $w' = m^2$. The expected responses may be calculated with a pooled slope determined as $b = T_b/2$ for a two-dose assay or as $b = T_b/4$ for a three-dose assay and with the unweighted means of the values of $\log \overline{y}'$ for each preparation. With these the expected responses Y are readily obtained. The antilog of each Y is m, leading to the working weight $w' = m^2$ and to the working response y by Eq. 74. The two dosage-response curves and their combined slope are computed with the coded doses and the values of w' and y, as are the χ^2 tests for the agreement of the slope of the unknown with that of the standard and of the observations with the computed lines. M and s_M are determined in coded dose units, the latter after converting w' to w. As a final step, both M and s_M are changed from coded doses to log doses by multiplying by i.

Example 19. The calculation of potency may be illustrated by a three-dose vitamin A assay of fish oil concentrate (U) in terms of the Canadian standard cod liver oil (S) from the reaction time of ovariectomized rats (48). The low dose of both the standard and the unknown was 51 known or assumed I.U. of vitamin A, each dose being multiplied by 1.5 to obtain the next larger dose ($i = 0.176$). The reaction time in days (y') of six

rats at each of the six dosage levels is shown in the upper part of Table XLVII.

The first step was to determine the mean reaction time in days $\bar{y}'$ and its logarithm (log $\bar{y}'$) at each dosage level. The standard deviation in terms of y' has been determined both from the sum of the squares $[y'^2]$ and from the range (R) at each dosage level, leading to $s = 1.1424$ and to the approximate $s = 0.9865$ respectively. Since the sum of squares within doses varied from 2.00 to 16.83, $\chi^2 = 7.326$ ($n = 5$, $P = 0.2$) was computed by Eq. 71, showing that the variances within doses could be considered as homogeneous and pooled.

TABLE XLVII. Assay of Vitamin A Content of a Fish Oil Concentrate (U) in Terms of the Canadian Standard Cod Liver Oil (S) by the Experimental Method in Table XLV, $S_1 = U_1 = 51$ known or assumed I.U. Vitamin A, Log-dose Interval $i = .176$

Data of Pugsley *et al.* (48)

	Reaction time in days (y') at dosage					
	S_1	S_2	S_3	U_1	U_2	U_3
	8	10	14	10	13	15
	10	11	15	9	13	16
	10	14	14	10	12	17
	8	12	14	10	11	15
	8	14	15	11	13	15
	11	10	16	10	12	17
$S(y')$	55	71	88	60	74	95
$\bar{y}'$	9.17	11.83	14.67	10.00	12.33	15.83
$y = \log \bar{y}'$	.962	1.073	1.166	1.000	1.091	1.199
$[y'^2]$	8.83	16.83	3.33	2.00	3.33	4.83
R	3	4	2	2	2	2

$S[y'^2] = 39.15$, $s_c^2 = 1.3050$, $n = 30$, $s_c = 1.1424$
$\bar{R} = 15/6 = 2.5$; approximate $s_c = 2 \times 2.5 \times .1973 = 0.9865$ (Table XXXII)
$T_a = .089$, $T_b = .403$, $i = .176$, $M' = .0518$ (Eq. 30)
$\bar{y} = 6.491/6 = 1.082$, antilog $\bar{y} = 12.08$
$s_M = 1.418 \times .176 s_c / .403\sqrt{6} \times 12.08 = .0209 s_c$ (Eq. 77)
With $s_c = 0.9865$, $s_M = 0.0206$; with $s_c = 1.142$, $s_M = 0.0239$

The shortened calculation of the log potency and its error with $y = \log \bar{y}'$ is given in Table XLVII. Both T_a and T_b were computed from the y's with the factorial coefficients in Table XXVII and substituted in Eq. 30 to obtain $M' = 0.0518$. To determine s_M by Eq. 77 required $\bar{y} = 1.082$ and its antilogarithm 12.08. Two estimates of the standard deviation in days s_c were available, $s_c = 0.9865$ as computed from the mean range, and the efficient estimate $s_c = 1.1424$ based upon the sums of squares within doses. These gave $s_M = 0.0206$ and 0.0239

TABLE XLVIII. Calculation of the Assay in Table XLVII by the Method of Maximum Likelihood

Preparation	Coded log dose x_1	log $\bar{y}'$	Expected response Y	Antilog Y m	Working weight $w' = m^2$	Working response y	Intermediate products $w'x$	$w'y$
S	0	.962	.966	9.25	86	.962	0	82.732
	1	1.073	1.067	11.67	136	1.073	136	145.928
	2	1.166	1.168	14.72	217	1.167	434	253.239
	Total	3.201			439		570	481.899
	Mean	1.067					1.298405	1.097720
U	0	1.000	.996	9.91	98	1.000	0	98.000
	1	1.091	1.097	12.50	156	1.091	156	170.196
	2	1.199	1.198	15.78	249	1.199	498	298.551
	Total	3.290			503		654	566.747
	Mean	1.097					1.300199	1.126734

$$w = 6 \times 2.303^2 w'/1.305 = 24.385 w'; \; i = .176$$

	S	U	Both
$[w'x_1^2]$	263.909	301.670	565.579
$[w'x_1y]$	26.7056	30.4140	57.1196
b	.10119	.10082	.100993
B^2	2.70240	3.06630	5.76869
$[w'y^2]$	2.70867	3.07337	5.78204
$[w'y^2] - B^2$	.00627	.00707	.01335
χ^2	.153	.172	.326

Converted from w' to w

$B^2 = 140.670$; $B^2 - t^2 = 136.828$

$S(w)_s = 10705$; $S(w)_u = 12266$

$s_M = .1332$ (Eq. 70); $M = .2855$ (Eq. 23a)

Corrected from coded doses to log dose

$M = .0502 \pm .0234$

respectively. Especially when the efficient estimate of s_c is used, the exact solution is seldom needed.

The calculation of the exact solution is exemplified in Table XLVIII. The expected responses Y have been computed from the coded $b = .403/4 = .10075$ and from the mean of the three values of log $\bar{y}'$ for the standard and for the unknown. These means were the "expected" values at $x = 1$, $Y_s = 1.067$ and $Y_u = 1.907$. Since $(x - \bar{x}) = -1$, 0, and 1 in the equation $Y = \bar{y} + b\ (x - \bar{x})$, the other values were obtained by adding b to and subtracting b from each mean. Their antilogs (m) were then squared to obtain the working weights (w') used in computing the intermediate products and the curves shown in the lower part of the table. Working responses were determined by Eq. 74. In this case they agreed almost exactly with the corresponding values of log $\bar{y}'$. The two values of χ^2 showed excellent agreement of the observations with the fitted lines, and no difference in slope. The log relative potency was then computed by Eq. 23a in coded log doses as $M = 0.2855$. To determine its error, B^2 and the sums of the weights were converted from w' to w by multiplying each by 24.385 leading by Eq. 70 to $s_M = 0.1332$. Both M and s_M in code were then multiplied by $i = 0.176$ to obtain the exact solution $M = 0.0502 \pm 0.0234$, substantially the same result as that obtained by the shortened method described above.

VII. Slope-Ratio Assays

In the preceding assays some unit of the response could be plotted as a straight line against the logarithm of the dose over an adequate dosage range. Parallel lines were fitted to concurrent tests with the standard and the unknown and the relative potency of the unknown determined on the log dose scale from the horizontal distance between them for a "log ratio" assay. This procedure has been illustrated by numerical examples of both animal and microbiological assays.

Sometimes, however, the response is a linear function of the dose rather than of the log dose. The tooth structure assay of vitamin C in Fig. 27,A (p. 257), for example, followed this pattern. The logarithm of the mean days cured in a bradycardia assay of vitamin B_1 (50) could be handled similarly, except that the dosage-response curves at zero dose showed a cure of 2.15 days, an improbable result. The most important applications, however, are in the microbiological assays where in many cases the response can be plotted linearly against arithmetic dosage units at the lower vitamin concentrations. Concurrent tests on two preparations which differ only in their content of a single active ingredient are fitted by two straight lines that converge at 0 dose. The potency of an

unknown relative to the standard can then be determined from the ratio of the slopes of the fitted lines. Assays meeting these conditions have been designated by Wood (51) as "slope-ratio assays." The term has since been extended (7) to assays where the response can be plotted linearly against some power of the dose. This may rectify those curves which cannot be plotted as straight lines against either the dose or the log dose. The present discussion, however, will be restricted to arithmetic doses, where the dose has an exponent of one. There are several advantages in following a balanced design and computing the results statistically. The discussion will be directed specifically to microbiological assays where at the present time the slope-ratio technique seems to have its greatest potential value.

1. Microbiological Assays and the Slope-Ratio Technique

In most microbiological assays the response is measured at several concentrations of the reference standard and of each of the one or more unknowns. The response may be measured turbidimetrically, in terms of pH, by titration, by weight, or by volume. The results for 8 or 10 doses of standard are then plotted on cross-section paper against the dose of vitamin, and a smooth curve is fitted by inspection. The response observed in each of several tubes of the unknown is referred to this line and its content of vitamin interpolated from the dosage scale on the abscissa. The amount of vitamin so interpolated is divided by the quantity of unknown originally placed in the tube to obtain the results in units per cubic centimeter. If the different estimates of potency agree within 10%, their average is accepted as the vitamin content of the unknown preparation. Except for this tolerance limit, there is no test of the precision of the assay. This common procedure (52) has several limitations which it is desirable to avoid.

A. Requirements for an efficient assay

One of the first requirements for an efficient microbiological assay is a linear relation between some function of dosage and response over an effective range. A change in medium may convert a curve that is linear against arithmetic dosage units to one that is linear against log dose. Thus in the riboflavin assay with *Lactobacillus casei* in the Snell-Strong medium, the titer plotted directly against the dose of riboflavin gave a straight line in the range up to 0.13 μg. of riboflavin. The dosage-response curve for the same organism in the Roberts-Snell medium gave a straight line over the same range when the log titer was plotted against the log dose (53). With the first medium a slope-ratio assay would be

suitable and with the second medium a log-ratio assay. With either unit, one could avoid the subjective factor due to interpolation from a free-hand curve of uncertain form.

The relation of response to arithmetic dose is often linear at the lower doses. When the concentration is increased sufficiently, the observed response falls below an extension of the linear part of the curve. This curvature may be methodological rather than inherent, so that the linear portion can be extended by a suitable change in technique. It may be due to an inadequate supply of some essential component in the basal medium which does not permit full utilization of the larger amounts of the vitamin. Doede (54) has shown, for example, that a change in the pH of the culture medium may extend the linear portion of the curve. Emery *et al.* (55) found a linear relation for inositol with one test organism and not with another. In some cases, such as in biotin (55) and in riboflavin assays (56), curvature has been observed at very small doses but a linear relation over an adequate range above 0.0001 μg. of biotin (*C. guillermondia*) and above .03 μg. of riboflavin (*L. helveticus*). Wood (56) has proposed adding sufficient vitamin to the basal medium in such cases to shift the entire assay into the linear portion of the curve, a procedure which does not affect the computation of the result in any way. These and other changes in laboratory technique may well increase the range of doses over which the response plots linearly against arithmetic dosage units. At the present time, however, the frequent occurrence of a slight convex curvature in the dosage-response curve is a major handicap in the wider adoption of slope-ratio or log-ratio microbiological assays.

In microbiological as in other assays, the experimenter assumes that the unknown preparation differs from the standard only by a dilution factor. This assumption should be tested quantitatively. If the concentration of the unknown could be adjusted to exact equality with the standard, their curves should coincide. When only the curve for the standard is plotted, there is no test for the similarity of the curve for the unknown. If the unknown preparation differs from the standard only in potency, their respective curves should be parallel in a log-ratio assay or converge at 0 dose in a slope-ratio assay. Wood (51) has found several slope-ratio assays in which the two lines did not converge at 0 dose, that for the unknown showing the higher response. This he attributed to nutritional deficiences in the basal medium which were supplied by the unknown in addition to its content of the vitamin under assay. With the usual technique such a qualitative difference can lead to an overestimate of the potency, though it may be detected by a "drift" in the interpolated potencies for successive doses of the unknown.

The replicates in an assay should contribute to the estimate of the assay error. In setting up a microbiological assay, duplicate or triplicate tubes are often prepared and handled together from the time the dose of vitamin is added until the titration is completed. In consequence, the variation between duplicates may underestimate the assay error with a discrepancy as much as tenfold (see example 10, pages 503-504). Environmental factors and an observer's bias in preparing and reading the tubes are probably responsible. The discrepancy could be avoided by arranging each assay in randomized groups. If complete randomization is not feasible, the more nearly a randomized arrangement is approximated in the preparation of the tubes, their handling and the reading of the results, the more nearly the variation in replicate tubes will indicate the error of the assay.

Large differences between the assumed and the assayed potencies of an unknown should be avoided. In such cases, the dosage-response curves are of very unequal slope. The variability about the curves of little slope may be considerably smaller than that about those of steeper slope. An error variance should not combine heterogeneous sources of variance. Hence assays in which the results largely confirm the assumption used in preparing the tubes are generally more reliable than those in which the assumed and observed potencies differ widely.

B. THE DESIGN OF BALANCED ASSAYS

If the experimental data meet the above requirements for a slope-ratio assay, the results can be evaluated statistically without other change in technique by the general procedure described by Finney (57). The calculation can be simplified, however, by the adoption of a suitable design, such as one of those described by Wood and Finney (58). They find little advantage in distributing the tubes unequally between the different dosage levels or between standard and the unknown, when a single unknown is assayed against the standard. The most efficient design is one using only a 0 dose and one relatively large dose of standard and of unknown. Such a three-point assay, however, has no test of its internal validity and cannot be recommended. Their common-zero 5-point assay has nearly as great an efficiency and tests convergence at 0 dose but not the linearity of the dosage-response curve. The following 4-dose or 9-point design for microbiological slope-ratio assays has the further advantage of testing the assumption of linearity.

From accumulated data on the standard a range of doses is determined within which the response plots linearly against arithmetic dosage units. No concentration of either standard or unknown should exceed the upper limit or "maximum." If there is a lower limit as well, enough

number, k, of dosage levels, coded as $x = 1, 2, 3, \cdots k$. Each randomized group, representing a replicate, may include two sets on the standard and at the zero dose to increase the reliability of the standard curve with which each unknown is compared. The first step in the calculation is to sum the responses in equivalent tubes. The sum (T_d) of the f responses (y) in equivalent tubes then represents r replicates of each dose of the $m - 1$ unknowns in the assay, h replicates of each dose of standard, and h' replicates of each tube of the negative control, totaling N tubes over the entire assay. The total of the responses for the control, for the standard and for all $(m - 1)$ unknowns are then added to obtain $S(y)$. The sum of the responses at each dosage level of each preparation (T_d) is multiplied by its coded dose (x) and the products summed to obtain $S(xT_d)_i = T_i$ for the standard and for each of the unknowns, identified by the subscripts 1 to m. The total of these sums of products for the standard and for all unknowns is designated as $S(T_i)$.

Given the above restrictions, it can be shown (59) that the computed response at $x = 0$ is

$$a' = \frac{2(2k+1)S(y) - 6S(T_i)}{N(k-1) + 3h'(k+1)} \tag{79}$$

If each preparation is identified by subscript i, its slope may be computed as

$$b_i = \frac{3}{2k+1}\left\{\frac{2T_i}{fk(k+1)} - a'\right\} \tag{80}$$

where $f = h$ for the standard $(i = 1)$, and $f = r$ for each unknown $(i = 2$ to $m)$. The following identity is true,

$$2Na' + hk(k+1)b_1 + rk(k+1)(b_2 + \cdots + b_m) = 2S(y) \tag{81}$$

The calculation should be checked in this way before proceeding further.

The potency of an unknown (i) relative to that of the standard (1) is computed in terms of coded doses as

$$J'_i = \frac{b_i}{b_1} \tag{82}$$

Each J' is converted to original units by multiplying it by the ratio of the actual dosage intervals I_1/I_i, the potency being

Example 20a. A "common-zero 5-point assay" for the riboflavin content of malt has been given by Wood (56), in which there were only two dosage levels ($k = 2$) but four tubes at each level ($f = h' = h = r = 4$), giving a total of $N = 20$ tubes. Although this design uses each observation more efficiently than distributing them among a larger number of dosage levels, it assumes that the relation between dosage and response is linear. The procedure in Eqs. 79 and 80 leads to the same answer as that described by Wood with no more labor and has greater flexibility. The data are shown in Table XLIX, there being no sep-

TABLE XLIX. A Common-Zero, 5-point Assay of the Riboflavin Content of Malt; Response Measured as Titer in milliliters 0.1 *N* NaOH

Data of Wood (56)

Preparation	Dose x	Individual titers, y				Total T_d	$S(y)$	$S(xT_d)$
Control	0	1.90	2.25	2.00	2.20	8.35	8.35	
Riboflavin	1	4.85	5.00	5.25	4.90	20.00		
	2	8.35	8.20	7.95	7.80	32.30	52.30	$T_1 = 84.60$
Malt	1	4.00	4.40	4.50	4.10	17.00	41.45	$T_2 = 65.90$
	2	6.05	6.20	6.10	6.10	24.45		
Total						102.10	102.10	$S(T_1) = 150.50$

$$a' = \frac{1021.0 - 903.0}{20 + 36} = 2.10714 \quad \text{(Eq. 79)}, \quad b_1 = \frac{T_1 - 12a'}{20} = 2.96572,$$

$$b_2 = 2.03072 \quad \text{(Eq. 80)}$$

aration of replicates into groups. The replicates have been totaled at each dosage level and again for each preparation. By Eqs. 79 and 80, $a' = 2.1071$, $b_1 = 2.9657$ and $b_2 = 2.0307$. The calculation was checked by sustitution in Eq. 81, giving the equality $40(2.1071) + 24(2.9657) + 24(2.0307) = 2(102.10)$. The potency in coded doses was computed by Eq. 82 as $J_2' = 2.0307/2.9657 = .6847$. The interval between successive doses of standard riboflavin was $I_1 = 0.10$ μg. and that for the unknown was $I_2 = 0.025$ g. of malt, giving $I_1/I_2 = 4$ μg. per gram. The riboflavin content of the malt, therefore, was $4 \times .6847 = 2.739$ μg. per gram.

B. THE CALCULATION OF POTENCY FROM A FOUR-DOSE ASSAY

When four dosage levels are used as in the proposed design for a balanced assay, the above equations may be solved with $k = 4$, leading to

giving $J = 0.449$, 0.288, and 0.301 μg. per milliliter of plant extract for preparations 2, 3, and 4 respectively. The computed dosage-response lines are shown in Fig. 8.

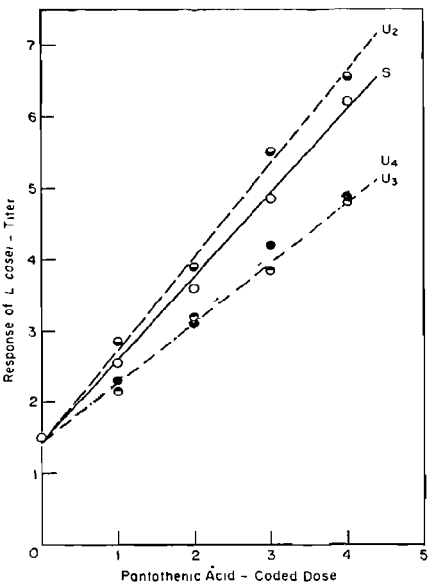

FIG. 8. Slope-ratio microbiological assay of the pantothenic acid content of plant tissues from data in Table LI.

C. ANALYSIS OF VARIANCE FOR SLOPE-RATIO ASSAYS

The first stage in the analysis of a slope-ratio assay is graphic. The individual responses or the dose means (or totals) are plotted on cross-section paper against the coded doses. Laying the edge of a transparent straightedge over the points is usually a sufficient test of their linearity. If at the top of the diagram they curve down, the tubes at the highest dosage levels may be omitted in computing the potencies. This preliminary inspection also tests whether the standard and the unknowns converge at zero dose, the requirement for qualitative equivalence in a slope-ratio assay. When the data are reasonably consistent, this graphic analysis will answer most purposes. Where curvature is suspected or the data are otherwise irregular, an analysis of variance can be used to test the validity of the assay.

The form of the analysis depends upon the design of the experiment. If it has been arranged in two or more independent groups which dupli-

cate one another, the differences between group totals are segregated as in any experiment in randomized groups. The discrepancy of the zero controls tests the linearity between dosage and response at the lower end of the curve. The comparability of the standard and the unknowns depends upon whether their dosage-response lines converge at zero dose within the sampling error. If three or more dosage levels have been used, the linearity of the lines can be tested either as a group or individually with the aid of orthogonal coefficients, or from the variation of dose means about the fitted lines. The remaining sum of squares provides an error for testing the significance of the terms which have been isolated. If they are mutually consistent, the test components are usually pooled with the remainder in estimating the error variance.

The sums of squares in a complete analysis may be computed with the work form in Table LII. The preferred design separates the two or

TABLE LII. Work Form for Computing the Sums of Squares in the Analysis of Slope-Ratio Assays

Row	Term	D.F.	Sum of squares
1	Replicated groups	$r - 1$	$rS(T_g^2)/N - C = [y^2]_g$
2	Slopes for m lines	m	$E - C$, where $E = a'S(y) + S(b_iT_i)$
3	Discrepancy of control	1	$S^2(y_0)/h' + E' - E$
4	Non-convergence at $x = 0$	$m - 1$	$S(C_i) + S(B_i^2) - E'$
5	Simple curvature	m	$S(Q_i^2)$
6	Scatter of dose means	$m(k - 3)$	$S(T_d^2/f) - C$ — (sums in rows 2 to 5)
7	Remainder or error	$N - mk - r$	By difference
8	Total	$N - 1$	$S(y^2) - C = [y^2]$
9	Correction for mean	1	$S^2(y)/N = C$

more replicates into r distinct groups of identical composition, a group containing at each dose one response of the $m - 1$ "unknown" preparations, h/r matched responses on the standard, and $h' = h$ or r negative controls. If replicates have not been separated into groups at the start, row 1 is omitted from the analysis. E in row 2 is the total variation in y accounted for by Eq. 78, as computed from all the data. Of this, C from row 9 represents the general mean, and the remainder, $E - C$, measures the effect of m slopes with a common origin at $x = 0$. E' in row 3 is obtained by recomputing a' and the b_i's without the negative controls, $S(y_0)$, and redetermining the equation for E in row 2 with these new values, as if $h' = 0$. When the same data are fitted with m independent lines, each with its own intercept and slope, the sum of squares for each line is $C_i + B_i^2$, defined as C and B^2 in Table IV (p.

463). The difference, $S(C_i + B_i^2) - E'$, with $m - 1$ degrees of freedom measures in row 4 the extent of non-convergence. In balanced assays, the sums of squares in rows 3 and 4 may be computed alternatively with the factorial coefficients described by Finney (76).

Simple curvature in row 5 is measured with the orthogonal quadratic coefficients x_2, computing Q_i^2 for each curve by Eq. 11. It tests whether the mean response can be related to the dose as satisfactorily by straight lines as by a series of simple curves. If the mean square is significant, some dosage-response curves may be linear and others curved. In this case Q_i^2 may be examined for each preparation, each with one degree of freedom. One or two in a series of unknowns may be responsible for the curvature and omitted from the main calculation. Sometimes the largest one or two doses may exceed the linear portion of the dosage-response curve, and in this case the assay may be recomputed, omitting the doses above the linear range. Alternatively, the experimental technique may need to be modified. More complex curvature or excessive scatter about the dosage-response lines is tested in row 6. If it is statistically significant in comparison with the error term in row 7 and if the replicate tubes were not randomized, the "excessive" variation may be due instead to handling duplicate tubes together during the assay. Whenever the mean square in row 6 is significantly greater than that in row 7, the sums of squares in rows 3 to 6 inclusive, if mutually compatible, are totaled to obtain an error variance for the assay.

TABLE LIII. Analysis of Variance for the Assay in Table XLIX Based upon the Work Form in Table LII

Term	D.F.	Sum of squares	Mean square	
Effect of b_1 and b_2	2	78.643	39.322	
Discrepancy of negative control	1	.006	.006	.17
Convergence of dosage-response lines	1	.085	.085	2.35
Remainder or error	15	.541	.0361	1.00
Total	19	79.275		
Correction for mean	1	521.220		
Composite assay error	17	.632	$.0372 = s^2$	

Example 20b. The variation in the riboflavin assay in Table XLIX could be separated into four components. The effect of the slopes in the first line of Table LIII was computed from a', b_1 and b_2 based upon all the data. For this purpose 5 decimal places in a' and b were required for 3-decimal accuracy in the analysis of variance. Omitting the negative controls, $S(y) = 93.75$ and $a' = 2.15625$, $b_1 = 2.93625$, and $b_2 = 2.00125$,

giving $E' = 582.438$. The discrepancy at zero dose was then computed as $8.35^2/4 + 582.438 - 599.863 = .006$. For testing their convergence at zero dose, the two dosage-response lines were computed separately. The sum of squares corresponding to the means of the two independently fitted lines were $(52.30^2 + 41.45^2)/8 = 556.674$. To obtain that for their slopes, the total response on the standard at the low dose was subtracted from the response at the high dose to obtain $S(x_1T_d)_1 = 12.30$. A similar calculation for the unknown gave $S(x_1T_d)_2 = 7.45$, from which $S(B_i^2) = (12.30^2 + 7.45^2)/8 = 25.849$. Hence the sum of squares for the four constants representing two independent straight lines for standard and unknown was the total of these components or 582.523. From this the sum of squares for three constants (a' in common) or $E' = 582.438$ was subtracted to obtain .085 as the sum of squares with 1 degree of freedom for testing the convergence of the two lines at zero dose. The total sum of squares was computed next as 79.275 and the error term with 15 degrees of freedom obtained as a difference. Neither the second nor third term exceeded the error significantly, so that the potency of the unknown could be computed from all the assay data and a composite error from the total variation about the fitted lines as $s^2 = .0372$.

TABLE LIV. Analysis of Variance of Assay in Table L

Term	D.F.	Sum of squares	Mean square	F
Effect of b_1 and b_2	2	110.742	55.371	
Discrepancy of negative control	1	.209	.209	1.91
Convergence of dosage-response lines	1	.084	.084	.77
Average curvature	2	.323	.1615	1.47
Scatter of dose means about curves	2	.924	.4620	4.22
Remainder or error	9	.986	.1096	1.00
Total	17	113.268		
Correction for mean	1	397.338		
Composite assay error	15	2.526	$.1684 = s^2$	

Example 21b. The analysis of variance in Table LIV for the riboflavin assay in Table L paralleled the preceding example for the first three lines which were computed similarly. With more than two dosage levels, the agreement with the hypothesis of linearity could also be tested. The coefficients x_2 for four doses were applied to the dose totals, T_d, for the standard and unknown separately, with $S(x_2T_d)$ equal to 1.07 and 1.20 respectively. $S(Q_i^2)$ was then determined as $(1.07^2 + 1.20^2)/8 = .323$ and the remaining curvature obtained as a difference from the variation in T_d. None of the 4 critical values exceeded the mean square for the remainder or error significantly, although the scatter of the dose means

approached the 5% level of significance. Because the tubes were known to have been randomized, the sums of squares in the last five lines have been totaled to obtain a composite assay error of $s^2 = .1684$.

Example 22b. The analysis of variance of the multiple pantothenic acid assay is given in Table LV. The intermediate steps in computing the sums of squares by the work form in Table LII are given with the original data in Table LI. Since none of the F values was significant, the sums of squares for all of the variation about the fitted lines in Fig. 8 have been totaled to obtain a composite error variance of $s^2 = 0.0295$ with 29 degrees of freedom.

TABLE LV. Analysis of Variance of Assay in Table LI

Term	D.F.	Sum of squares	Mean square	F
Effect of all slopes	4	65.840	16.460	
Discrepancy of negative control	1	.018	.018	.62
Non-convergence at 0 dose	3	.054	.0180	.62
Simple curvature	4	.061	.0153	.53
Scatter of dose means about curves	4	.228	.0570	1.96
Remainder from duplicates	17	.495	.0291	1.00
Total	33	66.696		
Correction for mean	1	511.694		
Composite assay error	29	.856	.0295 = s^2	

D. STANDARD ERROR OF POTENCY

The error variance for an assay (s^2) is based upon all the variation about the fitted lines and hence includes the components isolated in the analysis of variance with the exception of the sum of squares between randomized groups $[y^2]_g$. If the experiment is to be calculated as a slope-ratio assay, the other components should agree within the sampling error. The assay error may be computed without the analysis of variance as

$$s^2 = \frac{S(y^2) - E - [y^2]_g}{N - m - N_g} \tag{83}$$

where N_g is the number of randomized groups and E and $[y^2]_g$ are defined in Table LII. If differences between complete replicates or groups have not been segregated by the design, $[y^2]_g$ is omitted from Eq. 83, and the degrees of freedom in the denominator are increased to $N - m - 1$.

The standard error of the potency from a slope-ratio assay has been given in a general form (57, 58); for balanced slope ratio assays it may be written (59) as

$$s_{J'}^2 = \frac{6s^2}{(2k+1)b_1^2}\left\{\frac{h+rJ'^2}{rhk(k+1)} + \frac{3(1-J')^2}{N(k-1)+3h'(k+1)}\right\} \quad (84)$$

where all symbols have the same meaning as before. When the design has four dosage levels of the standard and of each unknown, the standard error takes the form

$$s_{J'}^2 = \frac{2s^2}{3b_1^2}\left\{\frac{h+rJ'^2}{20rh} + \frac{(1-J')^2}{N+5h'}\right\} \quad (84a)$$

To convert the standard error from relative to actual potencies, $s_{J'}$ is multiplied by the ratio of the dosage intervals, I_s/I_u.

Approximate confidence limits are given by $J \pm ts_J$ where t is obtained from the table of t with the degrees of freedom in s^2. Exact confidence limits for slope-ratio assays have been reported by both Finney (57) and Bliss (59), but in most microbiological assays the slope is so highly significant that the exact limits are nearly identical with their approximate values.

Example 20c. For the two-dose riboflavin assay in Table XLIX with 20 tubes, $s^2 = .0372$ and

$$s_{J'} = \sqrt{\frac{6\times.0372}{5(2.9657)^2}\left\{\frac{4+4(.6847)^2}{4\times4\times6} + \frac{3(1-.6847)^2}{20+36}\right\}} = .0184$$

by Eq. 84. This was multiplied by $I_s/I_u = 4$ to obtain s_J, so that the riboflavin content of the malt was assayed as 2.739 ± 0.074 μg. per gram.

Example 21c. The four-dose riboflavin assay with 18 tubes in Tables L and LIV had a larger error variance ($s^2 = .1684$), which accounted for its greater standard error by Eq. 84a of

$$s_{J'} = \sqrt{\frac{2\times.1684}{3(1.8853)^2}\left\{\frac{2+2(1.0015)^2}{20\times2\times2} + \frac{(1-1.0015)^2}{18+10}\right\}} = .0398$$

When multiplied by $I_s/I_u = 80$, the standard error of potency was 3.18 μg. per gram, the unknown having a riboflavin content of 80.12 ± 3.18 μg. per gram.

Example 22c. Since there were three unknowns in the pantothenic acid assay, it was convenient to multiply the standard error of relative potency in Eq. 84a by the ratio of the dosage intervals (0.4) to obtain

$$s_J = .02410\sqrt{\frac{1+J'^2}{10} + \frac{(1-J')^2}{11}}$$

Solving with the values of J' in Table LI, the standard error of potency was 0.0115, 0.0100, and 0.0102 μg. of vitamin per milliliter for plant extracts 2, 3 and 4 respectively.

VIII. Multiple or Repeated Assays

In the earlier sections of this chapter we have been concerned with self-contained assays from which the experimenter could compute the potency of an unknown and its sampling error without other data. When an assay procedure is used repeatedly in a given laboratory, an astute investigator is often able to spot results which seem abnormal and should be repeated. He draws on his past experience to interpret a current result. This is one phase of the more general problem of utilizing the data of other experiments in the analysis of a given biological assay or of analyzing jointly the results of several assays. We wish to use this wider experience quantitatively, in order to reduce the standard error of an estimated potency. The problem will be considered under three headings: the combination of individual assays of a single unknown into an improved estimate, statistical quality control in routine vitamin assays, and the design and analysis of collaborative experiments.

1. The Combination of Independent Assays of a Single Unknown

In order to obtain a more reliable estimate of its potency, an unknown preparation may be assayed independently at different times in the same laboratory or in several different laboratories. The problem arises as to how best to combine these data into a single estimate. Two procedures are possible. One is to work directly with the individual log potencies and their errors, M and s_M. If the potencies are computed directly, as in a slope-ratio assay, the potency (J) and 1 plus its proportionate standard error ($1 + s_J/J$) are transformed first to logarithms. How the log potencies are combined will depend largely upon whether or not they form a homogeneous series. The other method is to examine the homogeneity of the standard deviation and slope used in determining each M and s_M and, if consistent, to recompute each individual assay before combining them into a composite value. When applicable, more accurate results would be expected from the second method. The demonstration of a significant heterogeneity in the M's, of course, poses the question of finding and removing its source, a problem which is experimental rather than statistical.

A. The combination of homogeneous log potencies

As has been noted before, an individual standard error s_M is based upon the variation within the assay. An assay may be biased by factors which do not increase its standard error, such, for example, as an unequal loss in vitamin content in the preliminary handling of an aliquot

of the unknown. The magnitude of these factors cannot be predicted a priori, but their composite effect is measurable from the agreement of the individual M's. If the individual s_M's agree so closely upon inspection that the experimenter is willing to accept the k log potencies M in the series as equally reliable, their unweighted average ($\overline{M}$) may give a sufficiently accurate estimate of the potency of the unknown. The standard error of the mean potency $\overline{M}$ from such a series can be estimated directly from the k values of M as

$$\text{Unweighted } s_{\overline{M}} = \sqrt{\frac{S(M^2) - S^2(M)/k}{k(k-1)}} \tag{85}$$

with $k - 1$ degrees of freedom. Alternatively, the nearly equal standard errors, k in number, may be averaged and divided by $\sqrt{k}$ to obtain

$$\text{Crude } s_{\overline{M}} = \frac{S(s_M)}{k\sqrt{k}} \tag{85a}$$

Eq. 85 is usually preferred when $k \geqslant 8$.

The larger of these two estimates of $s_{\overline{M}}$ should be used whenever they differ by more than 20 to 30%. When they are about equal, one might average them. If the unweighted $s_{\overline{M}}$ is no larger than the crude $s_{\overline{M}}$, the only recognizable components of variation are those affecting each individual determination at its observed level of precision. By increasing sufficiently the number of observations in each separate assay, its standard error can be reduced to a level where relatively minor differences between assays appear as significant. When the unweighted $s_{\overline{M}}$ is consistently the larger, a more reliable mean log potency ($\overline{M}$) is obtained with a given number of observations by increasing the number of independent assays and using fewer observations in each one.

A weighted average of the M's is required when the standard errors s_M differ materially and average ten or more degrees of freedom. If based upon fewer degrees of freedom, the standard errors will not lead to sufficiently reliable individual weights. Under these conditions an unweighted average of the M's or a partially-weighted mean $\bar{M}$ may be used. With partial weighting about one-half of the log potencies, those with the smallest s_M's, are all given the same weight to keep a very few determinations with abnormally small standard errors from dominating the mean. With either partial or full weighting, the computation of the weighted mean log potency depends upon whether the individual M's of unequal precision differ more than would be expected by chance.

To determine whether the M's agree, each is assigned an observed weight inversely proportional to the square of its observed error or

$$W_1 = \frac{1}{s_M^2} \tag{86}$$

If partial weights are required, about one-half of the M's will have the same weight W_1', determined as

$$W_1' = S(n)/S(ns_M^2) \tag{86a}$$

The weighted sum of squares of the k individual assays about their weighted mean is defined as

$$[W_1M^2] = S(W_1M^2) - \frac{S^2(W_1M)}{S(W_1)} \tag{87}$$

If these weights were based upon the expected or true variances of the individual log potencies, $[W_1M^2]$ would be distributed as χ^2 with $k - 1$ degrees of freedom (60). The true variances, however, are presumably unknown and must be replaced in Eq. 86 by their estimates s_M. In consequence, Cochran (60) has shown that the homogeneity of the k independent M's is measured by

$$\chi_M^2 = (k-1) + \sqrt{\frac{\bar{n}-4}{\bar{n}-1}}\left\{\frac{(\bar{n}-2)[W_1M^2]}{\bar{n}} - (k-1)\right\} \tag{88}$$

where $\bar{n}$ is the average number of degrees of freedom in each s_M and preferably is constant from one assay to another in the series. When s_M is a large sample approximation with $n = \infty$, as in all-or-none assays computed with probits, $[W_1M^2]$ will be an approximate χ_M^2 with $k - 1$ degrees of freedom. If the different estimates of potency are mutually consistent within the precision of the respective s_M's, χ_M^2 should not exceed the value given in the χ^2 table for $P = 0.05$.

If the χ_M^2 test shows that the individual M's are homogeneous, the weighted mean log potency for the series is computed as

$$\bar{M} = \frac{S(W_1M)}{S(W_1)} \tag{89}$$

The precision of this weighted mean depends upon both the variance s_M^2 of each component and the variation between the individual log potencies. When the M's are in good agreement by χ^2, the variation between them is usually neglected. If each M and its error have been determined independently of the others, the standard error of the weighted $\bar{M}$ from Eq. 89 may be computed to a good first approximation * as

* I am indebted to W. G. Cochran for Eqs. 90 and 94 and for his aid in this section generally.

$$s_{\bar{M}} = \sqrt{\frac{\bar{n}\,[k(\bar{n}-2)+8]}{(\bar{n}-2)\,[k(\bar{n}-4)+12]\,S(W_1)}} \qquad (90)$$

where $\bar{n}$ is the average number of degrees of freedom in the individual estimates of s_M and is equal to 8 or more. The degrees of freedom for $s_{\bar{M}}$ is intermediate between $\bar{n}$ and $k\bar{n}$ and depends upon the variability in the weights. It may be computed as

$$\text{D.F. in } s_{\bar{M}} = \frac{\bar{n}S^2(s_M{}^2)}{S(s_M{}^4)} \qquad (91)$$

Example 23. The combination of homogeneous log potencies may be illustrated with the data reported by Irwin (27) from a collaborative growth assay of U.S.P. Reference Oil against the International Standard for vitamin A. Seven laboratories reported data with determinate fiducial limits on male rats for a five-week test period. The individual values of M and s_M as computed by Irwin are shown in Table LVI, the standard errors differing from those in Eq. 32 by omitting s^2t^2 from the

TABLE LVI. Collaborative Vitamin A Rat Growth Assay of U.S.P. Reference Oil (U) against International Standard (S), Based on Increase in Weight of Male Rats during a Five-Week Test Period as Reported by J. O. Irwin (27)

Lab. No.	M	s_M	$s_M{}^2$	W_1	W_1M	n
2	— .0689	.0620	.003844	260	— 17.9140	42
3	.0094	.0778	.006053	165	1.5510	42
5	.0652	.0515	.002652	377	24.5804	14
6	.0904	.0732	.005358	187	16.9048	20
7	.1166	.0924	.008538	117	13.6422	28
8	— .0393	.0805	.006480	154	— 6.0522	23
10	.0762	.0335	.001122	891	67.8942	50
Total			.034047	2151	100.6064	219

$[W_1M^2] = 11.3818 - 4.7056 = 6.6762$ (Eq. 87)

$\bar{n} = 219/7 = 31.286$; $\sqrt{(\bar{n}-4)/(\bar{n}-1)} = .9492$, $k = 7$

$\chi_M{}^2 = 6 + .9492(29.286 \times 6.6762/31.286 - 6) = 6.237$ (Eq. 88)

$\bar{M} = 100.6064/2151 = .0468$ (Eq. 89)

$$s_{\bar{M}} = \sqrt{\frac{31.286 \times 213.002}{29.286 \times 203.002 \times 2151}} = 0228 \quad \text{(Eq. 90)}$$

D.F. in $s_{\bar{M}} = 31.286 \times .0011592/.0002033 = 178$ (Eq. 91)

last term in the denominator. In view of the rather considerable variation in s_M, a weight W_1 was computed from each $s_M{}^2$ by Eq. 86 and the column W_1M written into the table. This led directly to $[W_1M^2] = 6.6762$ by Eq. 87. The degrees of freedom (n) in each s_M in the last column of the table were averaged to obtain $\bar{n} = 31.29$, with which by Eq. 88 $\chi_M{}^2 = 6.237$ with $k - 1 = 6$ degrees of freedom. In view of the

good homogeneity represented by this value, the mean log potency and its standard error ($\overline{M} \pm s_{\overline{M}}$) have been computed with Eqs. 89 and 90 as $\overline{M} = 0.0468 \pm 0.0228$. For determining confidence limits, the degrees of freedom in $s_{\overline{M}}$ were estimated by Eq. 91 as $31.286(.034047)^2/.0002033 = 178$.

B. THE COMBINATION OF HETEROGENEOUS LOG POTENCIES

When the individual M's, k in number, do not agree with one another by the χ_M^2 test, the variation between them cannot be neglected. The variance for each M then has two components, the variance internal to an assay and the variance between assays. The first component is measured by s_M^2, which may vary significantly from one assay to another. The second component will be designated as s_m^2 and is assumed to be the same for each M. As before, one would prefer to use the expected variances, but from the nature of the data these must be replaced by their estimates. Each M is assigned a semi-weight

$$W_2 = \frac{1}{s_M^2 + s_m^2} \tag{92}$$

Each s_M^2 is computed from the internal evidence of an individual assay as already defined. As Cochran (60) has shown, an unbiased s_m^2 may be computed from the *unweighted* values of M and s_M by the equation

$$s_m^2 = \frac{S(M^2) - S^2(M)/k}{k - 1} - \frac{S(s_M^2)}{k} \tag{93}$$

The semi-weights W_2 are more nearly equal than the W_1's defined in Eq. 86, especially if s_m^2 is considerably larger than any value of s_M^2. The mean log potency is recomputed by Eq. 89 substituting W_2 for W_1. The standard error of the semi-weighted $\overline{M}$ computed with the W_2's may be estimated as

$$s_{\overline{M}} = \sqrt{\frac{1}{S(W_2)}} \tag{94}$$

Example 24. Nine all-or-none assays based upon the incidence of rat fertility have been reported by Harris *et al.* (61) comparing natural α-tocopherol (U) with the synthetic preparation (S). Each assay was computed separately with probits as described in a preceding section of this chapter. The initial observations in terms of probits agreed with the hypotheses basic to the calculation, so that for the entire series $\chi^2 = 19.35$ with $n = 25$. The log potency (M) and its variance (s_M^2) are given in Table LVII for each assay. Using W_1 (Eq. 86) as the weight, $[W_1M^2] = 32.64$ with 8 degrees of freedom. In computing s_M^2, the ob-

TABLE LVII. Determinations of Vitamin E Potency of Natural α-Tocopherol (U) in Terms of Synthetic α-Tocopherol from Percentage Fertility in Rats; Each Assay Computed Separately with Probits

Data of Harris *et al.* (61)

Assay No.	M	s_M^2	W_1 (Eq. 86)	W_1M	$s_M^2+s_m^2$	W_2 (Eq. 92)
1	.041	.00171	585	23.985	.01123	89.0
2	.132	.00327	306	40.392	.01279	78.2
3	.112	.00569	176	19.712	.01521	65.7
4[a]	.210	.00480	208	43.680	.01432	69.8
5[a]	.096	.00308	325	31.200	.01260	79.4
6	.173	.00231	433	74.909	.01183	84.5
7	.174	.00204	490	85.260	.01156	86.5
8	.347	.00299	334	115.898	.01251	79.9
9	.384	.00493	203	77.952	.01445	69.2
Total	1.669	.03082	3060	512.988		702.2

[a] Single-dose assays; computed with combined slope from assays 3 and 6.

$[W_1M^2] = 32.637 \approx \chi_M^2, \quad n = 8$ (Eq. 87). $k = 9$

$s_m^2 = .01294 - .00342 = .00952$ (Eq. 93)

$\overline{M} = 127.58/702.2 = .1817$ (Eq. 89 solved with W_2)

$s_{\overline{M}} = \sqrt{1/702.2} = .0377$ (Eq. 94)

served ratios of χ^2/n have been replaced by one, their expected value, so that theoretically $n = \infty$ for each s_M^2. In consequence $[W_1M^2]$ itself is equivalent to χ_M^2, and it is evident that the individual assays differed much more from one another than would be expected from the inherent precision of the assay technique. Since much of the heterogeneity is due to the discordant values for assays 8 and 9, the experimenter usually would reexamine the data for these assays to determine whether they really belonged in the same series with assays 1 to 7 before proceeding with the calculation. We will assume here that he decided to retain them. The unweighted variance of M and the mean of the s_M^2's were substituted in Eq. 93 to obtain $s_m^2 = .00952$, with which semi-weights W_2 were determined for each assay. These varied much less than the W_1's and gave the semi-weighted mean $\overline{M} = .1817$ by solving Eq. 89 with the W_2's. The standard error of the mean log potency was computed with Eq. 94 as $s_{\overline{M}} = .0377$. Hence the potency of the natural α-tocopherol was assayed as $1.520 \pm .132$ that of the synthetic standard.

Example 25a. A sample of crystalline vitamin D_3 has been compared with the U.S.P. Reference Cod Liver Oil by Waddell and Kennedy (62) from tibia ash in a series of A.O.A.C. chick assays. The log potency and its variance as computed independently from the data of the first seven valid assays are given in Table LVIII, omitting test 107 in which

the slopes of the standard and unknown diverged significantly. To increase the degrees of freedom, the error variance for each assay included that measuring divergence in slope of the dosage response curves for standard and unknown.

The error variances were still based upon fewer than 10 degrees of freedom, so that partial weighting was preferred. Hence the four assays with the smallest s_M^2's, Nos. 98, 105, 109 and 110, were given the same weight $W_1' = 840$ as computed by Eq. 86a. By Eqs. 87 and 88, $\chi_M^2 = 10.278$ with 6 degrees of freedom, for which $P = 0.11$. Although $\overline{M} \pm s_{\overline{M}}$ might be computed by Eqs. 89 and 90 as $.7400 \pm .0213$, there were fewer degrees of freedom in $\overline{n}$ than required for Eq. 90 and χ_M^2 fell in the doubtful region approaching significance. For these reasons one

TABLE LVIII. Potency of Crystalline Vitamin D_3 in A.O.A.C. Units ($\times 10^{-7}$) as Determined by Chick Assays Computed Independently
Data of Waddell and Kennedy (62)

Assay No.	M	s_M^2	W_1	W_1M	n	W_2
98	.756	.001616	840	635.040	9	142
105	.695	.001183	840	583.800	6	151
106	.845	.003564	281	237.445	6	111
108	.904	.009604	104	94.016	6	66
109	.674	.001459	840	566.160	7	145
110	.808	.000384	840	678.720	7	171
111	.666	.003102	322	214.452	7	117
Total	5.348	.020912	4067	3009.633	48	903

$[W_1M^2] = 17.1177$ (Eq. 87); $\overline{n} = 6.857$, $\sqrt{(\overline{n}-4)/(\overline{n}-1)} = .6984$, $k = 7$
$\chi_M^2 = 10.278$, $n = 6$ (Eq. 88)
$s_m^2 = .008438 - .002987 = .005451$ (Eq. 93)
$\overline{M} = 679.576/903 = .7526$ (Eq. 89 solved with W_2)
$s_{\overline{M}} = \sqrt{1/903} = .0333$ (Eq. 94)

might prefer to use the semi-weights W_2. By Eq. 93 the variance between assays was $s_m^2 = .005451$, which in some tests was larger and in others smaller than the s_M^2 within assays. With the smaller variation between the W_2's (Table LVIII), the contribution of the seven assays to the mean was more nearly equalized, giving $\overline{M} = 0.753 \pm 0.033$ when Eqs. 89 and 94 were solved with W_2 as shown beneath the table.

C. THE COMBINATION OF ASSAYS COMPUTED WITH A COMMON ERROR VARIANCE AND SLOPE

Better estimates of the mean log potency have been obtained (63) when the individual M's were not computed independently. This re-

quires testing the stability over the series of assays of the error variance and the slope, which enter into the calculation of the log potency and its error.

The error variance s^2 in individual assays may have so few degrees of freedom as to increase appreciably the standard error of M as defined by either Eq. 90 or 94. If these separate estimates agree within their sampling error, each may be weighted by its degrees of freedom n to obtain a single, more reliable estimate of s^2 applying equally to all component assays. The homogeneity of s^2 is tested most readily by χ^2 as defined in Eq. 71. If homogeneous, the pooled s_o^2 for the group of assays with $n_o = S(n)$ degrees of freedom is used in computing each s_M^2.

A second statistic which may have a common value throughout a series of experiments is the slope b of the dosage-response curve. In calculating each individual assay the discrepancy between the slope of the standard and that of the unknown is tested before using the combined or assay slope. The next step is to test the agreement of individual assays in respect to their combined slopes. If b does not differ more between assays than would be expected from its sampling error, all the information on the slope can be pooled into a more reliable improved estimate.

If the error variance s^2 has proved stable over the series of experiments, the composite slope for all assays is computed by the second form of Eq. 24. The variation accounted for by this combined slope is B_c^2 as computed by Eq. 25, and the sum of squares for the variation in assay slope $[y^2]_b$ is determined by Eq. 26a, where B^2 is based upon the combined slope of each component assay. The mean square for $[y^2]_b$ with $k - 1$ degrees of freedom is divided by the combined error variance s_o^2 to obtain the variance ratio F for testing the homogeneity in slope. If individual assays are mutually consistent in this respect, their combined slope should lead to a more reliable estimate of the log potency M of each component assay than that determined entirely from its internal evidence. The pooled slope is especially useful where the individual assays would otherwise have indeterminate standard errors.

Granted homogeneity in these two basic statistics, one would prefer to use them in computing the individual log potencies M and their standard errors s_M. The assays in a series may differ in potency even though they agree in both their error variances and the slope of their dosage-response curves. Differences in their standard errors are then due only to the difference in the mean response, $\bar{y}_s - \bar{y}_u$, and in the number of observations on the standard and on the unknown, N_s and N_u. These factors control any differences in the weights W_1 used in comput-

ing the weighted mean log potency with Eq. 89. The agreement of the individual M's with this weighted mean is tested by χ^2 as defined in Eqs. 87 and 88. If χ^2 shows that the separate estimates agree within the sampling error, the standard error of the weighted mean may be computed by Eq. 90, replacing $\bar{n}$ with n, the degrees of freedom in the composite s_c^2 for the entire series. Should the M's differ significantly, $s_{\bar{M}}$ must include a factor representing the variability between the recomputed M's. As a rough approximation the discrepancy may be corrected by computing $s_{\bar{M}}$ as

$$\text{Approximate } s_{\bar{M}} = \sqrt{\frac{[W_1 M^2]}{S(W_1)(k-1)}} \tag{95}$$

without change in the weighted mean $\bar{M}$. For a more accurate estimate, one may compute a new series of weights W_2 as defined in Eqs. 92 and 93. These are used to obtain an improved estimate of the mean log potency and its standard error (Eq. 94), replacing $\bar{n}$ by the total degrees of freedom in s_c^2.

Example 25b. The vitamin D assays in Table LVIII have been reanalyzed as a group rather than individually. The error variances in the separate assays s^2 are shown in the second column of Table LIX, the degrees of freedom for each individual s^2 varying from 6 to 9. To test their homogeneity, each has been converted to its logarithm $(+1)$ for substitution in Eq. 71. Each s^2 was weighted by n in computing the error variance for the series as $s_c^2 = S(ns^2)/n_c = .8841$ with 48 degrees of freedom. The resulting $\chi^2 = 8.73$ with six degrees of freedom gave

TABLE LIX. Test for Homogeneity of the Assay Variance (s^2) and Slope (b) in the Chick Assays of Table LVIII

Assay No.	Assay s^2	n	$1 +$ log s^2	Combined $S + U$ $[xy]$	$[x^2]$	Slope b	B^2
98	.7737	9	.8886	6.8856	.5296	13.00	89.523
105	.5902	6	.7710	5.3249	.3463	15.38	81.879
106	1.4063	6	1.1481	5.8357	.4111	14.20	82.840
108	1.6535	6	1.2184	4.0468	.4108	9.85	39.865
109	.7363	7	.8670	8.8175	.6196	14.23	125.481
110	.1588	7	.2008	5.8481	.4407	13.27	77.604
111	1.0440	7	1.0187	7.1984	.6196	11.62	83.630
Total		48		43.9570	3.3777		580.822
Mean	.8841	→	.9465			13.014	

$s_c^2 = S(ns^2)/S(n) = 42.4370/48 = .8841$; $C' = 1 + (1.03970 - .02083)/18 = 1.0566$ (Eq. 72)

$\chi^2 = 2.180(45.4320 - 41.4279) = 8.729$, $n = 6$, $P = 0.19$ (Eq. 71)

$B_c^2 = (43.957)^2/3.3777 = 572.051$ (Eq. 25); $[y^2]_b = 580.822 - 572.051 = 8.771$ (Eq. 26a)

no evidence of heterogeneity which would prevent the use of the composite value. The terms required for comparing the assay slopes are shown in the right side of Table LIX. The individual slopes differed from $b = 9.85$ to $b = 15.38$ and the B^2 even more, the latter by Eq. 26a giving $[y^2]_b = 580.822 - 572.051 = 8.771$ with six degrees of freedom. These sums of squares have been entered in the analysis of variance of Table LX, which shows that, like the variances, the seven assay slopes could be replaced by their pooled value.

TABLE LX. Analysis of Variance for Homogeneity of Slopes in Table LIX

Term	D.F.	Sum of squares	Mean square	F
Combined slope, B_c^2	1	572.051		
Between assay slopes	6	8.771	1.4618	1.65
Error	48	42.437	$.8841 = s_c^2$	1.00

The log relative potency M of each assay was then recomputed with the combined slope $b_c = 13.014$ for the entire series and its error variance s_M^2 with the $s_c^2 = 0.8841$ with 48 degrees of freedom. The resulting values are shown in the second and third columns of Table LXI. The recomputed estimates agreed more closely with one another than the values in Table LVIII, as is evident from the comparative potencies in the last two columns of Table LXI. The weights W_1 applying to the

TABLE LXI. Comparison of Assays in Table LVIII Recomputed with the Combined Slope, $b_c = 13.014$ and Error Variance $s_c^2 = 0.8841$ with 48 degrees of freedom

Assay No.	M	s_M^2	W_1	W_1M	W_2	Potency ($\times 10^{-6}$) computed Separately	With b_c
98	.768	.001756	569	436.992	165	57.0	58.6
105	.712	.002368	422	300.464	150	49.6	51.5
106	.859	.002395	418	359.062	150	70.0	72.3
108	.851	.002391	418	355.718	150	80.2	71.0
109	.671	.002089	479	321.409	157	47.2	46.9
110	.809	.002182	458	370.522	155	64.3	64.4
111	.670	.002089	479	320.930	157	46.4	46.8
Total	5.340	.015270	3243	2465.097	1084		

$[W_1M^2] = 17.3406$ (Eq. 87), $\chi_M^2 = 10.388$, $n = 6$ (Eq. 88)

$\overline{M} = 2465.097/3243 = .7601$ (Eq. 89)

Approximate $s_{\overline{M}} = \sqrt{17.3406/(3243 \times 6)} = .0299$ (Eq. 95)

$s_m^2 = .006469 - .002181 = .004288$ (Eq. 93)

$S(W_2M) = 825.952$, $\overline{M} = .7619$ (Eq. 89 solved with W_2)

$n = 48$; $s_{\overline{M}} = \sqrt{1/1084} = .0304$ (Eq. 94)

recomputed estimates differed much less than before, from 418 to 569, but $\chi_M^2 = 10.39$, computed by Eq. 88, showed the same heterogeneity among the M's as before. The approximate standard error of the mean computed with the W_1's, $\overline{M} = 0.7601$, was determined by Eq. 95 to obtain $s_{\overline{M}} = .0299$. Alternatively, improved weights, W_2, were computed by Eq. 92, leading to a revised $\overline{M} = 0.7619 \pm .0304$. In original units the crystalline vitamin D_3 in these tests had a potency of $5.78(\pm 0.40) \times 10^7$ A.O.A.C. units of vitamin D. The values determined with the simpler weights, W_1, differed but little from those obtained with the W_2's. Although the mean log potency agreed closely with that obtained from individually computed assays, the reduction in its error was equivalent to increasing the number of assays by 20%.

2. Quality Control in Repeated Assays

The precision of the individual assays in a series can be increased if better estimates are available of the error variance s^2 and slope b When these statistics are stable over a period of time, one not only can combine the information in repeated assays of the same unknown, but can also draw on data accumulated in a given laboratory from earlier assays of other similar unknowns. This past experience can be utilized as described in the preceding section, comparing the slope and variance of each new assay with earlier composite values. If consistent, the laboratory values of s^2 and b should lead to an improved estimate of M and a smaller s_M.

Routine tests of homogeneity by χ^2 and by the analysis of variance, however, may not be practical. It is also possible that a progressive change in either the error variance or the slope might at first be overlooked without a continuous graphic check on their stability. Both purposes are served by quality control charts, in which a presumably stable variate is plotted against time. In the present case the variance s^2 may be plotted against the date on one chart and the slope b against the date on a second chart. A solid horizontal line corresponding to the average variance is drawn on the first chart and one corresponding to the combined slope b_c on the second chart. If the statistic is stable the plotted points will fall above and below each line at random. The lines are extended over future dates and as additional data are taken, they are plotted on the same charts and should fall similarly above and below the horizontal lines representing the laboratory means.

The mean line on a quality control chart is supplemented by parallel upper and lower dotted lines. These represent the limits of allow-

able variation, beyond which one would look for trouble. Any point which falls outside such a limit is potentially a significant deviation from the expected value defined by the solid central line. Although all points falling within the limits would be considered near enough the expected value to be in a state of "statistical control," a systematic trend toward either the upper or lower limit would be viewed with suspicion. The limits usually are spaced so that they enclose an expected percentage of the individual observations. Control charts (64, 65, 66) are often drawn with the so-called 3 σ limits, which are defined as "control limits." Three-sigma limits are based upon the normal distribution and would be expected to include approximately 99.7% of the population. The standard deviation (σ) is computed from the variability within the sample groups taken at successive intervals of time. In applying control chart procedure to biological assays, the limits will be based upon the tests of significance described earlier in this chapter.

Records which might otherwise be plotted in control chart form may prove unsuitable due to changes in the experimental design. Thus the error variance in successive assays may not be based on the same number of degrees of freedom. Since the standard error of a variance is a function of the degrees of freedom, the control limits are altered whenever these are changed. For simple and effective control charts, the variance from each unit assay should represent the same number of degrees of freedom. Similarly, the expected precision of the slope b should be uniform. This depends upon two factors, (a) "statistical control" in the error variance about the fitted lines and (b) the same combination of doses in successive assays so that $[x^2]$ is constant. A first requirement, therefore, in adopting a control chart technique is to stabilize the assay procedure in these respects. When s^2 (or s) and b are both "out of control," their ratio, $s/b = \lambda$, may prove stable, a condition which is of more limited value for improving assay precision.

A. CONTROL CHARTS FOR THE ERROR VARIANCE

In computing the results of most unit assays, we obtain initially the sum of the squared deviations from the dosage-response lines for standard and unknown. With a standardized procedure this sum of squares or "reduced $[y^2]$" in successive assays will be based upon the same number of degrees of freedom (n). The homogeneity of a series of such sums of squares can be tested in terms of their variances by χ^2 as defined in Eqs. 71 and 72. Alternatively, χ^2 may be computed directly from the k sums of squares when each is based upon the same n degrees of freedom as

$$\chi^2 = \frac{2.303 n_c}{C'} \left\{ \log \overline{[y^2]} - \frac{S(\log [y^2])}{k} \right\} \tag{71a}$$

where $\overline{[y^2]}$ is the arithmetic mean of the k individual $[y^2]$'s, $n_c = kn$ is the total number of degrees of freedom in the series and

$$C' = \frac{3n_c + k + 1}{3n_c} \tag{72a}$$

If χ^2 indicates a significant or near-significant heterogeneity due to a very few component assays, the most aberrant one or two may be omitted, so as to bring χ^2 nearer to $k - 1$, the degrees of freedom in χ^2. A control chart for future sums of squares ns_c^2 may be started with the data from a group of demonstrably homogeneous assays and the combined s_c^2 from the pooled sum of squares. Its purpose is to test whether new values of s^2 from individual assays are consistent with this composite older value.

The critical statistic for such a test is the variance ratio F. In the absence of a control chart, the s^2 from each new assay could be divided by the composite s_c^2 from past homogeneous assays to obtain the variance ratio F. If the degrees of freedom n were uniform in each future assay, its sum of squares could be tested directly by dividing by ns_c^2. In transferring this procedure to a control chart, the sum of squares in each individual assay is plotted on the ordinate against the date on which the assay is made on the abscissa. A horizontal line is drawn on the chart at ns_c^2, about which the observed values should fall in approximately equal numbers.

Above and below the "expected" value, two lines are drawn representing percentage limits based upon the F distribution and so spaced that they enclose 95% of the plotted points in statistical control. The percentage limits are computed with the aid of a table of F at the 2½% point (67). For the upper percentage limit the F value (F_U) is that with $n_2 = n_c$, the degrees of freedom in s_c^2, and $n_1 = n$, the degrees of freedom in the error sum of squares for each individual assay. For the lower limit, the F value (F_L) is that with n_1 and n_2 reversed, so that $n_1 = n_c$ and $n_2 = n$. The percentage limits are then defined as:

$$\text{Upper limit} = ns_c^2 F_U$$

and (96)

$$\text{Lower limit} = \frac{ns_c^2}{F_L}$$

These limits are narrowed as the degrees of freedom in s_c^2 increase with time. Until they are based upon many degrees of freedom, all three horizontal lines on the quality control chart should be recomputed fre-

quently. For a chart in terms of the error variance s^2 instead of the sum of squares, the limits given by Eq. 96 are divided by n.

With some assays, the standard deviation may be determined from the mean range rather than from the square root of the error variance. In these cases the quality control chart for testing the error variance is constructed in terms of the mean range rather than of the sum of squares $[y^2]$. For setting control limits the reader is referred to the papers by Lord (31) and by Grubbs and Weaver (30).

Example 26a. Control charts for vitamin assays may be illustrated with data on the vitamin D assay from the percentage tibia ash of chicks. These have kindly been made available by Dr. J. Waddell and Mr. G. H. Kennedy of the E. I. duPont de Nemours and Co. They cover a period of 28 months and in more than three-fourths of the assays doses of 4, 8 and 12 units of standard vitamin D were fed to duplicate cages of chicks. In each of the nine assays which varied from this pattern, the statistics other than the observed error variance and slope have been adjusted to make them comparable with the rest of the series. The essential statistics for each curve are given in Table LXII. The data from 17 assays accumulated through December, 1941, have been used to set up a control chart procedure and the results in 1942 used to check the stability of the earlier statistics.

The first stage in analyzing the error variance was to examine its homogeneity about the individual dosage-response curves. The sums of squares through 1941 in Table LXII, each with four degrees of freedom,

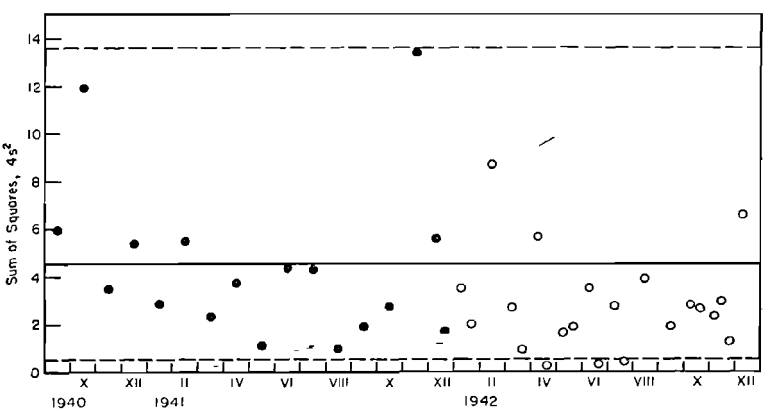

FIG. 9. Quality control chart for testing the stability of the error variance in Table LXII in terms of ns^2 (= $[y^2]$). Horizontal lines have been computed only from the data plotted as solid circles but extended through 1942.

TABLE LXII. Dosage-Response Curves for the Standard in Routine Chick Assays for Vitamin D from the Percentage Tibia Ash. The Sum of Squares for Each Assay, $[y^2]$, has Four Degrees of Freedom ($n = 4$) and Each Slope b Is Assumed to Have Been computed with $[x^2] = 0.2329$, so that $B^2 = 0.2329b^2$. Control Charts in Figs. 9 and 10 Determined with Data through December, 1941, and Tested with Results in 1942.

Data from J. Waddell and G. H. Kennedy, E. I. duPont de Nemours and Co.

Date 1940-1941	Sum of squares $[y^2] = ns^2$	Standard deviation s	Slope b	$s/b = \lambda$
Sept.	5.979	1.223	17.46	.0700
Oct.	11.882	1.724	18.04	.0956
Nov.	3.517	.938	19.75	.0475
Dec.	5.406	1.162	17.94	.0648
Jan.	2.880	.849	17.44	.0487
Feb.	5.487	1.171	14.87	.0787
Mar.	2.349	.766	9.68	.0791
Apr.	3.773	.971	17.21	.0564
May	1.131	.532	16.70	.0319
June	4.394	1.048	14.21	.0738
July	4.342	1.042	12.62	.0826
Aug.	1.012	.503	13.97	.0360
Sept.	1.929	.694	15.42	.0450
Oct.	2.777	.833	13.60	.0613
Nov.	13.400	1.830	18.75	.0976
Dec.	5.626	1.186	14.12	.0840
Dec.	1.752	.662	16.57	.0400
Total	77.636		268.35	

From all 38 ($= k$) curves:

$S(b^2) = 7323.55$ $[b^2] = 351.54$

$S(bs) = 473.654$ $[bs] = 17.964$

$S(s^2) = 34.2728$ $[s^2] = 4.4890$

Date 1942	Sum of squares $[y^2] = ns^2$	Standard deviation s	Slope b	$s/b = \lambda$
Jan.	3.561	.944	14.56	.0648
Jan.	2.030	.712	12.83	.0555
Feb.	8.712	1.476	14.40	.1025
Mar.	2.746	.828	12.86	.0644
Mar.	.971	.493	13.71	.0360
Apr.	5.690	1.192	14.10	.0845
Apr.	.284	.266	14.34	.0185
May	1.678	.648	13.71	.0473
May	1.916	.692	13.74	.0504
June	3.581	.946	9.50	.0996
June	.352	.297	9.44	.0315
July	2.799	.837	9.72	.0861
July	.444	.333	7.06	.0472
Aug.	3.947	.993	11.58	.0858
Sept.	1.929	.694	11.71	.0593
Oct.	2.836	.842	11.11	.0758
Oct.	2.694	.821	9.23	.0889
Nov.	2.381	.771	10.21	.0755
Nov.	3.002	.866	9.79	.0885
Nov.	1.308	.572	10.56	.0542
Dec.	6.602	1.285	12.21	.1052
Grand total		33.642	514.72	2.5145

have been plotted as solid circles against the date in Fig. 9. To test their homogeneity χ^2 was computed with Eq. 71a, converting each $[y^2]$ and their mean, $\overline{[y^2]} = 4.567$, to its logarithm. For the series of 17 curves

$$\chi^2 = \frac{2.303 \times 68}{1.088}\left\{.6596 - \frac{9.4669}{17}\right\} = 14.78 \text{ with 16 degrees of freedom.}$$

Since χ^2 was less than its degrees of freedom, the error variance in these 17 assays did not differ significantly from their mean of $s_c^2 = 1.1417$ with $4 \times 17 = 68$ degrees of freedom.

Since the individual assays were homogeneous through 1941, their mean sum of squares $ns_c^2 = 4 \times 1.1417 = 4.567$ determined the position of the solid line in the control chart of Fig. 9. To obtain percentage limits which would enclose 95% of future homogeneous assays, F was interpolated from a table of the variance ratio for the 2½% point (67). For $n_1 = 4$ and $n_2 = 68$, $F = 2.981$; for $n_1 = 68$ and $n_2 = 4$, $1/F = .1198$. Substituting these values in Eq. 96 the upper limit was determined as $4.567 \times 2.981 = 13.614$ and the lower limit as $4.567 \times .1198 = .547$. These have been drawn in the diagram as broken lines which enclosed all solid circles.

Both the solid line for the expected ns_c^2 and the percentage limits were then projected through 1942. Of the individual sums of squares in the following year, 18 out of 21 fell below the expected value and three of these below the lower percentage limit. It is apparent, therefore, that the variance was not stabilized at its original level but dropped in 1942 to a lower value. The error variance in 1942 of $s_c^2 = .7079$ with 84 degrees of freedom was significantly less than that through December 1941, confirming the conclusion from the control chart.

B. CONTROL CHART FOR SLOPE

Granted a stable error variance, the stability of the slope can be tested by a second quality control chart. The standard error of b as given in Eq. 13 is $s_b = s/\sqrt{[x^2]}$, which resembles the standard error of the mean of a normal distribution. The denominator of the error is also used in computing each individual slope and depends upon the number of doses used and their distribution. Hence it is subject to experimental control. The s in the numerator is the square root of the error variance (s_c^2) in the same assays; presumably s^2 has been shown to be stable over the given period. If this is indeed true, the stability of the b's in the same assays can be tested by the analysis of variance defined in Eqs. 25 to 27. For this purpose it is sometimes convenient to compute the effect of each slope as

$$B^2 = [x^2]b^2 \tag{9b}$$

and that of the combined slope as

$$B_c^2 = \frac{[x^2]S^2(b)}{k} \tag{25b}$$

Our objectives are to convert this general procedure to a control chart basis using the limits for sample means drawn from a normal population.

The limits for the slope could be computed as tolerance limits based upon the normal distribution. In this case limits would include a given proportion of the population with a preassigned probability. A good approximation for such tolerance limits could be computed from the mean slope, the error variance and values taken from tables of the χ^2 distribution and of the deviates of the normal distribution. The procedure is described in a paper by Wald and Wolfowitz (68).

For testing the agreement of a new slope with past experience, percentage or control limits are probably as suitable and easier to compute. The solid central horizontal line is that for b_c. When the error variance is stable and each slope is computed with the same denominator, this is equivalent to the arithmetic mean of the individual slopes or $b_c = \overline{b}$. The proposed limits may then be computed as

$$\text{Control limits} = b_c \pm ts_b \tag{97}$$

where Student's t is that for the degrees of freedom in the pooled variance s_c^2 over all assays at the required level of P, usually .05 and s_b is computed by Eq. 13 with the $[x^2]$ for each individual assay in the denominator. These control limits should lead to the same conclusion as an analysis of variance of the b's.

When the slope is not stable, past experience is of limited value. Gridgeman (21) has reported a marked secular trend in the slope of the dosage-response curve for the vitamin A assay which would preclude the use of a single laboratory slope. In order to reduce the sampling error of the slope for a single assay, he averaged the slopes from the two preceding assays with that observed in the current test. Whether this would be preferable to using the assay slope alone would need to be determined in each case.

Example 26b. The individual slopes for the assays in Table LXII have been plotted in Fig. 10. Since the error variances in these assays proved homogeneous, the slopes through December, 1941, (solid circles) have been tested for homogeneity by an analysis of variance. The effect of each slope was computed by Eq. 9b as $B^2 = .2329b^2$. Since $[x^2]$ was constant, the effect of the combined slope could be determined by Eq.

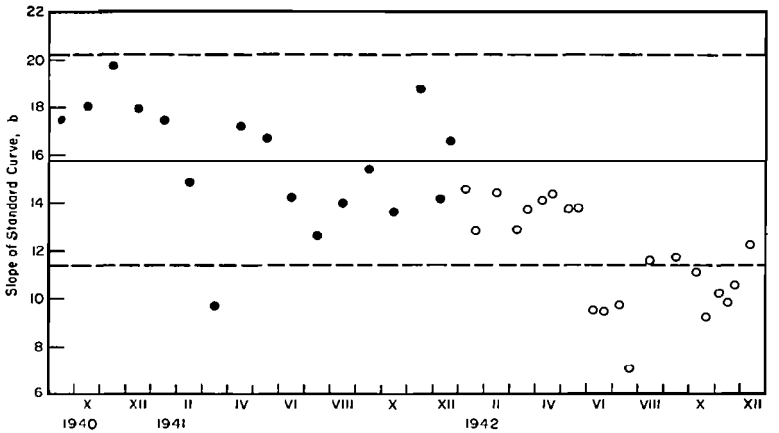

FIG. 10. Quality control chart for testing the stability of the slope b in Table LXII. Horizontal lines have been computed only from the data plotted as solid circles but extended through 1942.

25b as $B_o^2 = .2329\ (268.35)^2/17 = 986.561$ and the variation in b by Eq. 26a. These sums of squares gave the analysis of variance in Table LXIII in which the individual slopes were found to agree with their pooled value in 1940-41. Within this period the slope was in statistical "control," except for a suspicious but not significant downward trend.

TABLE LXIII. Analysis of Variance of the Slopes of the Assays through December, 1941, in Table LXII

Term	D.F.	Sum of squares	Mean square	F
Effect of the combined slope	1	986.561	986.561	
Variation in individual slopes	16	24.420	1.5262	1.34
Composite error	68	77.636	1.1417	1.00

The mean slope of $\overline{b} = 268.35/17 = 15.785$ was then drawn as a solid line in Fig. 10 and projected into the following year. Equation 13 gave the standard error of each individual slope as $s_b = \sqrt{1.1417/.2329} = 2.214$. For 68 degrees of freedom and $P = .05$, $t = 1.995$, so that the control limits by Eq. 97 were $15.785 \pm 4.417 = 20.20$ and 11.37, which have been drawn as broken lines in Fig. 10 and projected through 1942. In 1942 every individual slope fell below the expected value and 9 of the 21 values below the lower control limit. It is obvious that the slope in 1942 differed significantly from that in 1941.

C. CONTROL CHART FOR λ

The construction of control charts for the error variance and the slope is based on the assumption that s and b are independent of each other. If they should be correlated, use of a "laboratory" slope could reduce the precision of individual assays. The possibility remains, however, that the standard deviation in terms of the log dose might be stable despite variation in both the standard deviation of the response and the slope of the dosage-response curve. This is measured by $\lambda = s/b$ (Eq. 16). A stable λ would indicate that the assay was uniformly precise, even though relative potency could not be computed with a pooled value of the slope. The laboratory λ could be used, however, in determining s_M, and its many degrees of freedom would narrow the confidence limits of a given assay.

The presence of a significant relation between the standard deviation in y about the fitted lines (s) and the slope (b) can be determined by computing the variance ratio (F) for the regression of s upon b as

$$F = \frac{(k-2)[bs]^2}{[b^2][s^2] - [bs]^2} \tag{98}$$

The sums of squares and products in Eq. 98 are computed with Eq. 4, 5 and 8, substituting b for x and s for y. In testing the significance of an observed F, $n_1 = 1$ and $n_2 = k - 2$ where k is the number of assays or paired observations of s and b.

If s is related to b, the validity of a composite or laboratory λ_c should be checked. This requires that $s = \lambda b$ over the assays in the series. If s is plotted on the ordinate against b on the abscissa, the observations should agree with a straight line which passes through the origin at $s = 0$ and $b = 0$. Non-agreement with a zero intercept may be computed (69) as

$$[s^2]_0 = \frac{S^2(s)}{k} + \frac{[bs]^2}{[b^2]} - \frac{S^2(bs)}{S(b^2)} \tag{99}$$

The variance of $[s^2]_0$ may be computed as

$$V([s^2]_0) = \frac{[s^2] - [bs]^2/[b^2]}{k-2} \tag{100}$$

which is analagous to Eq. 10. F with $n_1 = 1$ and $n_2 = k - 2$ is computed by dividing $[s^2]_0$ by its variance.

If there is adequate agreement with a zero intercept, λ may be computed either by averaging the λ for all individual assays or as

$$\hat{\lambda} = \frac{S(bs)}{S(b^2)} \tag{101}$$

and the two estimates should not differ significantly. The variance of each λ with $k - 1$ degrees of freedom may be computed as

$$s_\lambda^2 = \frac{S(s^2) - S^2(bs)/S(b^2)}{\bar{b}^2(k-1)} \tag{102}$$

where $\bar{b} = S(b)/k$. This is an alternate to Eq. 18, which assumes that s and b are not correlated. It is the appropriate error when λ is determined from a series of assays and s and b are correlated. The variance of $\hat{\lambda}$ for the entire series of k assays may then be computed as

$$s_{\hat{\lambda}}^2 = \frac{s_\lambda^2}{k} \tag{103}$$

The control chart for λ is then prepared with the solid line for the expected value at $\hat{\lambda}$ and the percentage limits are computed as $\hat{\lambda} \pm ts_\lambda$. If the precision of the assay technique is "in control" and s is correlated with b, the assay slope must be used in computing M, but s_M may be determined with $\hat{\lambda}$. Equation 32 is modified to the form

$$s_M = \hat{\lambda}\sqrt{\frac{1}{N_u} + \frac{1}{N_s} + \frac{(\bar{y}_u - \bar{y}_s)^2}{B^2 - (\hat{\lambda}bt)^2}} \tag{32e}$$

where t and s_M have as many degrees of freedom as $\hat{\lambda}$. Other forms of the same equation (32a to 32d) may be changed similarly.

Example 26c. The corresponding drop in ns^2 and in b in the control charts for the chick assay for vitamin D in Figs. 9 and 10 suggests that s and b may not have been independent. In order to include the discrepancy between 1940-41 and 1942, the variance ratio for testing the significance of the regression of s upon b has been computed from the data for the entire period. The terms required for the calculation are shown below Table LXII and have been substituted in Eq. 98 to obtain

$$F = \frac{(38-2)(17.964)^2}{(351.54)(4.4890) - (17.964)^2} = 9.254$$

with $n_1 = 1$ and $n_2 = 38 - 2 = 36$. The probability of an F ratio of this magnitude arising from unrelated variables was determined from Table V in reference (1) as $P < .01$, which rejects the hypothesis that s was independent of b.

Since the standard deviation was related to the slope, s plotted against b might lead to a straight line passing through the intercept of $s = 0$ and $b = 0$. This could be tested by computing $[s^2]_0 = 33.642^2/38 + 17.964^2/351.54 - 473.654^2/7323.55 = .0680$ by Eq. 99 and its variance by Eq. 100 as $V([s^2]_0) = (4.4890 - 17.964^2/351.54)/(38 - 2) =$

0.0992. The ratio of these terms, $F = .0680/.0992 = 0.69$, was less than 1, indicating good agreement with a zero intercept. In consequence a stable value of λ could be estimated by Eq. 101 as $\hat{\lambda} = 473.65/7323.6 = .06468$. The error of λ for a single assay was determined from the observed variation by Eq. 102 as

$$s_\lambda = \sqrt{\frac{34.2728 - 473.654^2/7323.55}{13.545^2(38 - 1)}} = .0232$$

The individual λ's were then plotted against date in the control chart of Fig. 11. The solid line for the expected value of $\hat{\lambda} = 0.0647$ divided

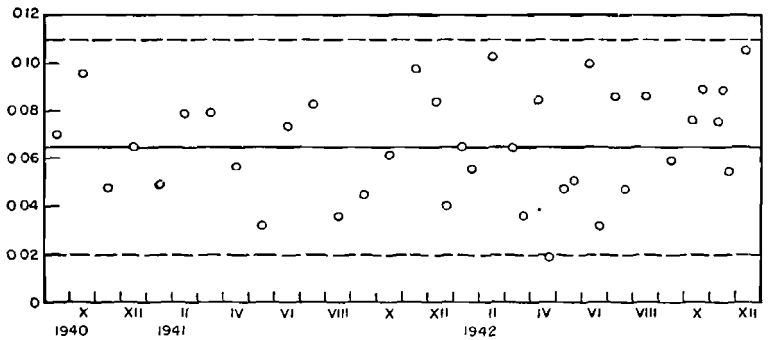

FIG. 11. Control chart for λ in Table LXII, with horizontal lines computed from values of s and $\bar{b}$ for the entire series of 38 assays.

the observations nearly equally, both through December, 1941, and in 1942. The percentage limits at $P = .05$ ($t = 2.027$) were $\hat{\lambda} = 0.0647 \pm .0450$ and enclosed 37 of the 38 observations when computed with the expected standard deviation. It is evident that the precision of the assay procedure in these experiments was in statistical control. The mean and standard deviation of λ have also been computed from the individual λ's directly. The mean $\bar{\lambda} = 2.5145/38 = .06617$ differed but little from that computed with Eq. 101, while the standard deviation of $s_\lambda = 0.0225$ differed by less than 3% from the estimate computed with Eqs. 102.

3. Collaborative Experiments

A favorite device for developing an assay technique is the colloborative experiment. Sometimes its primary purpose is instructional, to give experience with a method intended for official use in many laboratories. More often, however, the collaborative experiment is intended to answer questions such as: Do the collaborators agree with one another suf-

ficiently to warrant the introduction of the new method as outlined? Or, what is the potency of a new standard preparation and what is the precision of the estimate? If collaborative experiments are to fulfil these broader objectives, they must be well designed statistically and the original data should be analyzed by a competent statistician. Without these precautions, much valuable time and effort can be wasted in ill-conceived collaborative studies.

For a satisfactory collaborative experiment, the separate assays should meet the requirements of a good design as they have been defined in the preceding sections. Each component assay should be planned so that it will provide a quantitative estimate of potency and its error, M and s_M, with a minimum of calculation and high efficiency. Each collaborator should follow the same basic design in respect to number of doses, dosage interval, and the number of observations. Unless the procedure is known to be quite uniform, it is helpful to have the assay repeated in each laboratory on two or more different days. It is of especial importance for collaborators to follow instructions exactly, and for this purpose they need to understand the reasoning back of each requirement. Failure to understand the reasons for a given design is probably responsible for many of the lapses in collaborative studies.

In planning a collaborative experiment, it is desirable to have more than one "unknown" preparation assayed against the standard. At least one of the unknowns should be an undisclosed dilution of the standard or of another unknown. By this means, the results of each laboratory can be checked against an absolute criterion. Discrepant laboratories can be identified and their estimates of potency eliminated from the general average for each unknown.

An alternative or supplementary procedure is to distribute four or five different unknowns to each collaborator. The agreement of each laboratory with the overall average can then be compared for each unknown in the series. A collaborator deviating widely from the average in respect to several unknowns may then be scrutinized more closely and, if desirable, his results eliminated from all averages. In a U.S.P. collaborative study of penicillin (71), for example, the discrepancy of some laboratories proved to be due to an unusually strong or weak solution of the standard preparation, since the potency of the unknown penicillins relative to one another agreed well with the general average.

In analyzing the results of a collaborative experiment, the underlying statistics used in computing the M's and s_M's may be compared, especially the values for the slope b and error variance s^2. Sometimes these have proved to be stable between different laboratories so that pooled values could be used. The variation between the M's computed from

the data of each laboratory should be compared with that expected from their respective s_M's, using the methods described in a preceding section. In some cases they have been found to agree satisfactorily with one another, as in the collaborative vitamin A assay analyzed by Irwin (27). If they differ significantly from one another, further study of the method is desirable. Some assay characteristics which should be watched in analyzing a collaborative study have been outlined in collaborative studies of the chick assay of vitamin D (72, 73). Other papers reporting collaborative drug assays (74, 75) have described methods of statistical analysis that are applicable to experiments with vitamins.

Glossary of Symbols

There is no standard system of statistical symbols, although some conventions have wide acceptance. These and all other symbols are defined in the text when they first appear and as often afterwards as seems necessary. The text definitions take precedence over those in the glossary, should they differ. Because the number of letters is limited, some symbols have been used with different meanings in different sections. Conversely, different applications of what is basically the same statistic may be represented by different symbols. The appropriate meaning in any given computation should be clear from the context.

Although there are exceptions, Latin letters refer to observed values and the statistics computed from them and Greek letters to the unknown population parameters which are estimated by the statistics. Subscripts and superscripts often qualify the more general term to which they are attached but in other cases they merely extend the alphabet, as will be evident from the list. The references to equations, tables and text are illustrative only.

a — the position of the dosage-response curve at $\bar{x}$, equal numerically to $\bar{y}$ for all dosage levels included in the curve. Eq. 1.

a' — the zero intercept of the regression line or the expected reponse Y at $X = 0$. Eq. 1a.

A — any specified mean square in an analysis of variance.

b — the regression coefficient or slope of a straight dosage-response line, frequently qualified by identifying subscripts. Eq. 6.

b_o — the combined slope of two or more parallel dosage-response lines. Eq. 24.

b_i — slope of a dosage-response curve in a slope-ratio assay (Eq. 80), where $i = 1$ is the standard and $i = 2$ to m are the unknowns (Eq. 78).

b_s, b_u — slopes of dosage-response lines for standard and for unknown respectively.

b_v — slope of the dependent variate y upon a concomitant measure v in covariance.

b_1 — (1) slope of the dosage-response curve in terms of the coded log-dose x_1. Eq. 7a.
(2) slope of the dosage-response curve for the standard in a slope-ratio assay. Eq. 80.

b' — the partial regression coefficient for x^2 in a parabola.

B^2 — that part of the sum of squared deviations in y accounted for by the slope of the dosage-response line, frequently used with subscripts for identification. Eq. 9.

B_c^2 — B^2 for the combined slope. Eq. 25.

B_i^2 — B^2 for an individual curve, such as B_s^2 for the standard and B_u^2 for the unknown.

B_v^2 — B^2 for the slope of the dependent variate y upon a concomitant measure v in covariance.

C — correction for the mean in an analysis of variance. Table IV.

C^2 and $C = \sqrt{C^2}$ — correction for computing the exact confidence limits of M from its approximate limits. Eq. 35.

C' — correction used in computing χ^2 for testing the homogeneity of three or more variances. Eq. 72.

d_1, d_2 — intra-pair differences in response, $d_1 = S_2 - U_1$ and $d_2 = U_2 - S_1$, used in computing two-dose assays from paired observations.

D_1, D_2, D_3, D_4 — the initial differences in computing a two-dose factorial assay, defined on page 509.

D^2 — variance for the factorial comparison of Standard and Unknown, computed from T_a. Table XXII.

$D.F.$ — degrees of freedom. Table IV, Eq. 91.

E — correction for the mean and two or more slopes in

	the analysis of variance of a slope-ratio assay. Table LII.
E'	same as E but computed without the responses at zero-dose.
f	frequency or the number of observations at a given dose or in a given group.
F	variance ratio for testing whether two variances agree within the sampling error. Table IV.
F_h	variance ratio for testing the divergence of the dosage-response curves for the two unknowns in an assay of two unknowns from groups of three responses. Eq. 47.
F_k	variance ratio for testing the discrepancy in slope between the standard and the two unknowns in an assay of two unknowns from groups of three responses. Eq. 47a.
F_L, F_U	F values used in computing the lower and upper percentage limits for the variance in a quality control chart. Eq. 96.
g	number of unknown preparations in an assay when determining number of observations to be assigned to each unknown and the single standard. Page 484.
G	mean square for groups in an analysis of variance. Table X.
h	number of replicates at each dose of the standard in a slope-ratio assay.
h'	number of replicates of the negative control in a slope-ratio assay.
i	interval between equally-spaced successive log-doses or doses of vitamin. Eq. 29.
i	as a subscript refers to the individual components in a series.
I_i	interval between successive doses of unknown i, where $i = 2$ to m, in a slope-ratio assay. Eq. 82a.
I_1	interval between successive doses of the standard in a slope-ratio assay. Eq. 82a.
I'	the ratio of the interval i (in logarithms or other units) between successive doses of vitamin to the interval between the coded coefficients x_1.

J	potency of an unknown relative to that of the standard in original units. Eq. 82a.
J'_i	potency of unknown i relative to that of the standard in the coded units of a slope-ratio assay. Eq. 82.
k	(1) number of dosage levels of a given preparation. Eq. 2(b). (2) number of rows, columns and treatments in a Latin square. Table XIII. (3) number of groups in an all-or-none assay in which the expected number in the smaller class, either positive or negative, is equal to 0.5 individual or more. Eq. 63. (4) number of replicated assays in a series. Eq. 85.
K	coefficient for estimating the standard deviation from the mean range. Table XXXII and Eq. 48.
m	(1) antilogarithm of Y, the log-response expected at a given dose of vitamin A in a vaginal smear assay. Eq. 73. (2) number of different preparations (standard and unknowns) in a balanced slope-ratio assay. Eq. 78.
M	log-ratio of potency, corrected for the difference between the assumed dose of the unknown and the actual dose of the standard. Eq. 23a.
M'	log-ratio of potency of an unknown in units assumed equal to the standard. Eq. 23b.
n	degrees of freedom in a sum of squares, mean square. error variance, χ^2 or other statistic.
n_s, n_u	degrees of freedom in the error variance for the standard and unknown respectively. Eq. 27.
n_1, n_2	the degrees of freedom in the larger mean square and in the smaller mean square respectively for determining the significance of an observed F. Page 464.
n'	number of $k \times k$ Latin squares in an experiment with k rows and k treatments in common. Table XIV.
$\bar{n}$	average number of degrees of freedom in each standard error (s_M) of the replicated assays in a series. Eq. 88.
N	total number of observations. Eq. 2.
N_g	number of randomized groups in a slope-ratio assay. Eq. 83.

N_s, N_u total number of observations on the standard and on the unknown. Eq. 32.

N'_s, N'_u total number of animals in an all-or-none assay on doses of the standard and of the unknown having an expected response between 3.5 and 6.5 probits. Eq. 54.

p proportion observed to give a positive reaction in a single group of an all-or-none assay. Eq. 56.

p_1, p_2 individual pair totals in a factorial assay from paired observations, where p_1 is the sum of the responses to U_2 and S_1 and p_2 to U_1 and S_2, either single pair being designated as p in Eq. 41.

P (1) probability on a scale of 0 to 1 of the chance occurrence of a discrepancy as great as or greater than that observed in a test of significance.
(2) probability that the confidence limits computed from M and s_M will not bracket the true log-potency. Page 506.
(3) the expected proportionate response to a given dose in an all-or-none assay. Eqs. 55, 56.

P_1, P_2 sums of single-pair totals p_1 and p_2. Eq. 41.

q, Q $= 1 - p$ and $1 - P$ in computing curves from all-or-none data. Eq. 56.

Q^2, Q^2_t effect of quadratic curvature in the dosage-response curve in terms of y^2, measured over all preparations (Eq. 11) or for an individual preparation. Table LII.

r (1) number of replicates at each dose of the $m - 1$ unknowns in a slope-ratio assay. Eq. 80.
(2) correlation coefficient. Eq. 104.

R range or difference between the smallest and the largest number in a series. Page 510.

$\bar{R}$ mean of several similar ranges of series or subsets of equal length, where subscripts identify the series from which the range is determined. Table XXXII, Eq. 48, 48a.

s standard deviation.
(1) without subscripts, an estimate of the population σ in units of the individual variate y computed efficiently as $\sqrt{\bar{s}^2}$ or approximately from $\bar{R}$.
(2) with subscripts a, b, J, J', λ, M, $\bar{M}$, s, Y, the

standard deviation or standard error of the statistic designated by the subscript.

s_X, s_Y — standard error of the log-dose X computed for any given Y (Eq. 66) or of the expected response Y computed for any given X. Eq. 14.

s^2 — standard deviation squared or error variance,
(1) without subscripts, usually in units of the individual deviate squared and often computed as the mean square for error in an analysis of variance. See $s(1)$.
(2) with the subscripts in $s(2)$, the square of the standard errors or the error variances of the statistics in the subscript.

s_c^2 — a combined error variance. Eq. 96.

s_k, s_k^2 — inter-group standard deviation and error variance respectively. Eqs. 46a and 48a.

s_m^2 — the error variance among unweighted M's in a series of replicated assays. Eq. 93.

s_p^2 — inter-pair error variance. Eq. 41.

$[s^2]_0$ — variance for the non-agreement with a zero intercept of the line relating s and b from independent assays. Eq. 99.

S — the standard preparation in a biological assay, also used as a subscript with the same meaning.

S_1, S_2, S_3 — (1) a specified dosage level of the standard preparation.
(2) response to the specified dosage level. Page 509.

$S(\ )$ — sum of all values included in the parentheses, such as $S(y)$ is the sum of the y's; sometimes abbreviated to T or to T with a subscript.

S.E. — approximate standard error of relative potency. Eq. 33.

t — (1) value read from the Student distribution at a given probability (P) and degrees of freedom, n. Eqs. 18, 34.
(2) ratio of a difference to its standard error. Eq. 28.

T — total of the y's or $S(y)$, without subscripts summed over all categories.

T — with subscripts a, b, ab, c, ac, as in Tables XXII and

	XXIV, are the individual sums of products $S(xT_d)$ for factorial treatment totals, any one of them being designated as T_i.
T_c	column total in a Latin square (see alternate definition for T_c above).
T_d	dose total or total of the y's at each dose.
T_g	group total or total y in a group, such as a litter.
T_{iv}, T_{iy}	sum of products of V_d or Y_d with a given set i of factorial coefficients x. Eq. 49.
T_r	row total in a Latin square.
T_s, T_u, $T_{u'}$	total responses on the standard and on the unknowns U and U'. Table XXXV.
T'	without subscripts and with subscripts c, d, g, and r refer to totals in a balanced experiment from which a response metameter is missing and must be replaced by Eq. 21 or 22.
T'	with subscripts a, b, ab, c, ac, $(= i)$ stand for their respective factorial treatment totals after adjustment by covariance for variation in the concomitant measure v. Eq. 49.
U, U'	the unknown preparation or preparations in a biological assay, also used as a subscript with the same meaning.
U_1, U_2, U_3	(1) dosage level of the unknown preparation. (2) response to the specified dosage level. Page 509.
v	an initial or concomitant variate to be adjusted by covariance.
V_d	total of the v's for a given dose and preparation. Page 529.
$V(\ )$	indicates the error variance of the variate or statistic enclosed in parentheses.
w	weight given to each response when the variance of the response is a function of the dose. Eqs. 55, 73.
W_1	weight assigned to a log-potency M based upon its error variance s_M^2. Eq. 86.

W_1'	partial weight assigned to the more precise assays and based on their average s_M^2's. Eq. 86a.
W_2	semi-weight, assigned to each M in a series when the s_M's differ significantly. Eq. 92.
x	(1) the observed log-dose or dose of vitamin against which the response can be plotted as a straight line. (2) factorial coefficients representing doses of vitamin, used in isolating the treatment effects of a factorial assay.
x_s, x_u	observed log-dose of standard and unknown respectively.
x_1	coded log-dose (or dose) used in computing the slope of a straight line in a balanced assay or dosage-response curve. Page 458.
x_2	quadratic coefficients orthogonal with x_1 used in computing Q^2. Eq. 11.
X	(1) a contemplated log-dose used in solving a regression equation for a predicted Y. Eq. 1. (2) the log-dose predicted by the regression equation for a given reponse Y. Eq. 1c.
X_L, X_L'	confidence limits for (1) the log-ratio of potencies M or M'. Eqs. 34 and 34a. (2) the log dose predicted for a given Y. Eq. 67.
X_s, X_u	log-dose of standard and of unknown as predicted from their respective dosage-response curves for a given Y.
y	an observed response in the units (metameter) used in computing a dosage-response curve or assay.
y_a, y_b, y_{ab}	sums of the products of factorial coefficients x with the four y's within a randomized group of 4. Page 509.
y_d	mean response to a given dose.
y_h, y_j, y_k	differences and sum of the responses within groups in an assay of two unknowns in groups of three. Eq. 44.
$y_s, y_u, y_{u'}$	an observed response to the standard and to two unknown preparations respectively. Eq. 44.
y_1, y_2	response y on low and high doses respectively.
y'	(1) computed response to replace a missing observed y. Eq. 21 or 22.

	(2) reaction time of an individual rat in days in a vaginal smear vitamin A assay. Page 550.
$[y^2]_b$	sum of squares between slopes. Eq. 26.
$[y^2]_g$	sum of squares between randomized groups. Eq. 83.
Y	response predicted by the dosage-response equation for a contemplated log-dose X. Eq. 1.
Y_d	dose total or sum of the y's at a given dose of a single preparation of vitamin. Page 529.
Y_L	confidence limits of Y at a contemplated log-dose X. Eq. 15.
$Y_1, Y_2 \ldots Y_m$	response predicted by dosage-response curves for preparations 1, 2, . . . m in a slope-ratio assay. Eq. 78.
Z	ordinate of the normal curve at the expected proportionate response P, used in calculations of all-or-none data. Eqs. 55, 56.
a	population parameter estimated by a.
β	population parameter estimated by b.
λ	standard deviation s of a single response y converted to units of the log-dose x. Eq. 16.
$\hat{\lambda}$	estimated value of λ for a series of assays in statistical control. Eq. 101.
μ'	population log-potency estimated by M'. Eq. 105.
σ	population standard deviation estimated by s.
χ^2	a statistic used in testing the agreement of observed and expected responses or variances. Eqs. 63, 71.
χ_b^2	χ^2 for testing the significance of a difference in slope of two or more weighted regression equations. Eq. 68.
χ_M^2	χ^2 for testing the homogeneity of the M's in k independent assays. Eq. 88.
Assay χ^2	total variability in an assay computed with weights. Eq. 69.
[ ]	sum of the terms inside the brackets, expressed not in original values but as deviations from their means.
—	bar over a symbol denotes the mean, such as $\bar{x}$ is a mean of the x's, $\bar{M}$ is a mean of the M's.

Addendum

A. HOMOGENEITY OF THE ERROR FOR RANDOMIZED GROUPS

When experiments are arranged in randomized groups or blocks, the appropriate error for comparing treatment effects is usually the interaction of groups or blocks by treatments. In biological assays which follow this design, the mean square for the interaction of groups by doses has been used almost universally in estimating the error s_M of the log potency M. When M has been determined for a given unknown in two or more assays and the variation in M compared with the individual values of s_M, agreement has been common. When they have differed, s_M has underestimated the error, presumably because the estimates of relative potency depended also upon factors which differed between assays. The combining of such independent assays is discussed on pages 576–586.

In some recent experiments, however, the agreement among independent estimates of potency has been suspiciously better than would be expected from their respective s_M's. This has led Finney (77) to a more explicit formulation of the mathematical model underlying the assay. He assumes that the unknown is qualitatively identical with the standard, except for an undetermined amount of an inert diluent, and that the dosage-response curve is linear. Groups are presumed to vary in their expected slopes, which introduces a separate variance component for slope. This same component also affects the difference in the response of individual groups to the standard and to an unknown that differs in potency. Since it occurs in both the numerator and denominator of M, the variability in the log potency would not be changed by this component.

Although the variance component for slope, if it occurs, does not affect the estimation of potency, it increases the two interactions of groups by preparations and groups by slope. Its importance can be examined easily in factorial assays that are arranged in randomized groups. The factorial coefficients in the work forms for 2×2 (p. 494) and for 2×3 (p. 497) assays are applied separately to the data for each group. For two-dose assays, the three factorial effects may be isolated as described on pages 509–512 in terms of y_a, y_b, and y_{ab}. In a three-dose assay, the additional terms y_c and y_{ac} are required. From the mathematical model, the mean square among litters in y_{ab}, in $y_{c\bar{}}$ and in y_{ac} depends only upon the random variance σ^2. The corresponding variances from y_a and from y_b may contain in addition a variance component for differences from litter to litter in the expected slope, a component which cancels out in computing M. On this hypothesis, only the contributions to s^2 from y_{ab}, y_c, and y_{ac} should be used in computing λ and s_M.

This restriction would reduce markedly the degrees of freedom in the error variance, as Finney notes. For two-dose factorial assays of a single unknown, two-thirds of the degrees of freedom would be lost, and for three-dose assays, two-fifths. The fewer degrees of freedom in s^2 would increase the value of t used in computing the confidence interval for M. If the number of groups were small, this would tend to offset any reduction in s^2. Intra-group replication is seldom a practicable solution and would complicate the analysis. If the variability in y_a and in y_b were of the same magnitude as that of the remaining factorial effects, a small potential bias might be accepted in return for a more stable estimate.

Five experiments among the examples were arranged in randomized groups or litters, in each with a significant difference between litters. The appropriate mean squares have been computed from the y_a's, from the y_b's, and from the remaining treatment effects for each litter. The results are shown in Table LXIV. In no case was there a clear demonstration (at $P \leqslant .05$) of the variance component for slope implied by Finney's model. In Examples 8 and 13, the assayed log potency M' was very close to a true value of $\mu' = 0$. The variance component for slope in the expected mean square for y_a is multiplied by μ'^2, so that in these experiments the mean square for preparation should approach that for the remainder.

TABLE LXIV. Subdivision of the Error Variance (s^2) into Mean Squares for the Interaction of Litter with the Difference between Standard and Unknown, of Litter with Slope, and the Remainder

Example No.	M.S. for interaction with litter of						Correlation r of y_a and y_b
	n	Prep'n.	n	Slope	n	Remainder	
4	11	5.544	11	13.681	10 [a]	11.202	— .217
8	11	9.889	11	7.975	33	6.076	— .528
11	9	62.06	9	36.68	9	24.84	.153
13	11	3.280 [b]	11	4.741	22	2.556	.083
15	5	.00118	5	.00133	15	.00162	.187

[a] Reduced by missing value, replaced so as to minimize this interaction.
[b] Based upon the difference between the standard and both unknowns.

As a corollary of the model, a greater or lesser slope in a given litter should affect both y_a and y_b. If y_a and y_b were not correlated, the variance component for slope would not cancel out in computing M, and a greater variability in both terms would belong in s_M. This relation can be determined by calculating the correlation between y_a and y_b from the variances of T_a and T_b, and their covariance, as defined in Eq. 106 of the next section. The denominators of $V(T_a)$, $\mathrm{Cov}(T_a T_b)$, and $V(T_b)$ are constant and cancel out in computing the correlation coefficient,

$$r = \frac{\mathrm{Cov}(T_a T_b)}{\sqrt{V(T_a)\,V(T_b)}} \tag{104}$$

The significance of r, which may vary from -1 to 1, can be determined from Table VI in reference (1). The correlations for the five examples in Table LXIV are given in the last column. None was significant at $P \leqslant .05$, and two of the five were negatively correlated, where only positive correlations would be expected.

In view of these empirical findings, the applicability of the model may be questioned. It depends upon the assumption of qualitative identity which underlies an analytical assay (p. 450). In comparative assays of an unknown which differs chemically from the standard, its relative potency may vary with the species of test animal, as in the differential response of rats and chicks to Vitamins D_2 and D_3, or with the experimental conditions. If the assumption of qualitative identity is not warranted, some groups may react more strongly to one preparation than to another, even though the slopes of their dosage-response curves may be the same. This would introduce a variance component in the y_a's not present in Finney's model.

The dosage-response curve in most graded-response assays is assumed at best to be an elongated sigmoid with a substantially linear central section. Dosages are selected which are believed to fall within this linear zone. However, the response in some groups may overlap into the curved portion approaching an upper or a lower bound. This would add still another variance component, increasing the interaction of groups with parallelism and with curvature. It seems doubtful, therefore, in comparative assays that any of the interactions can be trusted to represent only the random variance σ^2.

The empirical evidence of the present assays suggests that the variance component for slope may often be negligible. This is supported by the agreement among independent determinations with the same unknown, which is rarely better than would be expected from their respective s_M's. For general purposes, the interaction of doses by groups seems to lead to a suitable estimate of the error variance s^2 for computing s_M, especially when the assayed potency of the unknown does not differ appreciably from the potency assumed when selecting the doses. However, when there is a marked difference in response to the standard and to the unknown, or when an assay seems to have an unduly large or unexpectedly small error, the subdivision of the interaction of groups by doses should be illuminating. These factorial constituents may be examined for homogeneity and for the correlation of the variation in y_a with that in y_b. If the variances in y_a and in y_b are significantly larger than the

remaining interactions and correlated with each other, either s^2 may be computed from the remaining interactions of treatments with groups, or exact confidence limits may be determined from the variance for y_a and for y_b and their covariance, as described in the next section. If the mean squares for these interactions differ but are not correlated, the assay error may be computed with two separate estimates of s^2 by the equations in the next section.

B. THE PRECISION OF AN ESTIMATED POTENCY WITH TWO ERROR VARIANCES

Heterogeneity in the error variance may arise either inadvertently from variance components which, unlike Finney's model, do not cancel one another, or intentionally from the nature of the design. If the assay represents a factorial experiment in randomized groups, confidence limits can be computed directly from the observed variances for the numerator and denominator of M and their covariance. Alternatively, it may be rational to assume a covariance of zero and compute the standard error and confidence limits with different error variances for numerator and denominator. These alternatives may be considered separately.

Confidence limits are calculated directly and without assumptions as to variance components with Fieller's fundamental formula (78). The log potency of a two-dose assay, for example, is determined from the ratio $M' = iT_a/T_b$ (Eq. 29), which is an estimate of an unknown true value $\mu' = \eta/\xi$. The numerator and denominator are assumed to be subject to random errors that are normally distributed and independent of η and ξ. These errors are estimated by the variances and covariance $i^2V(T_a)$, $i\mathrm{Cov}(T_aT_b)$, and $V(T_b)$, respectively, each with n degrees of freedom. Then it can be shown that the confidence limits for M' in a two-dose assay are the roots of the quadratic equation

$$\begin{aligned}\mu'^2\{T_b^2 - t^2V(T_b)\} - \mu' 2i\{T_aT_b - t^2\mathrm{Cov}(T_aT_b)\} \\ + i^2(T_a^2 - t^2V(T_a)\} = 0 \end{aligned} \qquad (105)$$

where t is read from a table of the Student distribution at the desired P and $n = f - 1$, and i is the interval between doses in logarithms. The required variances and covariances are computed from the f values of y_a and y_b, defined on page 509, to obtain

$$\begin{aligned} V(T_a) &= \{fS(y_a^2) - T_a^2\}/n \\ \mathrm{Cov}(T_aT_b) &= \{fS(y_ay_b) - T_aT_b\}/n \\ V(T_b) &= \{fS(y_b^2) - T_b^2\}/n \end{aligned} \qquad (106)$$

When these values are substituted in Eq. 105, it reduces to

$$A\mu'^2 + B\mu' + C = 0$$

from which the confidence limits are the familiar roots

$$X_L' = \frac{-B}{2A} \pm \frac{\sqrt{B^2 - 4AC}}{2A} \tag{107}$$

In a three-dose factorial assay, where $M' = 4iT_a/3T_b$ (Eq. 30), y_a and y_b are determined for each group with the factorial coefficients for T_a and T_b in Table XXIV. The variances and covariances of T_a and T_b are calculated as above with Eq. 106 and substituted in the following equation for the three-dose assay:

$$\mu'^2 9\{T_b^2 - t^2V(T_b)\} - \mu' 24i\{T_aT_b - t^2\mathrm{Cov}(T_aT_b)\} + 16i^2\{T_a^2 - t^2V(T_a)\} = 0 \tag{105a}$$

When reduced numerically, the confidence limits are the two values of μ' determined as its roots with Eq. 34b.

The mean difference in response and the slope may be determined from data differing in origin, so that they are not correlated. Thus, in a three-dose collaborative assay of Vitamin D with rats, litter mates were assigned to five different preparations at each dosage level and slope was determined from comparisons between litters. Hence, the variance for slope included an additional component for litters. An analagous situation arises when the slope is stable over a series of experiments, and the confidence limits are computed with more than one variance and t (28).

In units of a single observation, the error variance for the difference in the mean response on the standard and on the unknown may be designated as s_1^2 with n_1 degrees of freedom and that for slope as s_2^2 with n_2 degrees of freedom. Confidence limits in log units of assumed potency may then be computed by an extension of Eqs. 32 and 34 as

$$X_L' = C^2M' \pm \frac{C}{b_c}\sqrt{s_1^2t_1^2\left\{\frac{1}{N_s} + \frac{1}{N_u}\right\} + \frac{s_2^2t_2^2(\bar{y}_u - \bar{y}_s)^2}{B_c^2 - s_2^2t_2^2}} \tag{108}$$

where t_1 and t_2 are tabular values of the t-distribution for the required significance level with n_1 and n_2 degrees of freedom, respectively, and $C^2 = B_c^2/(B_c^2 - s_2^2t_2^2)$. By adding the logarithm of the assumed potency of the unknown to the midpoint of the interval, the confidence limits for the assayed potency may be determined in units of vitamin.

Example 11a. Confidence limits computed by Fieller's technique may be compared with those based upon two estimates of a single variance, one from the interaction of groups by parallelism and the other from the pooled error. The vitamin A growth assay in Example 11 (p. 511) has been used for this purpose. For assumption-free limits, the variances and covariances from Eq. 106 were $V(T_a) = \{10 \times 4699 -$

as 1502 units per gram within limits at $P = .05$ of 1156 and 1808 units per gram.

TABLE LXVI. Factorial Analysis, Estimated Potency, and Confidence Limits for Vitamin A Assay in Table LXV

Effect	Dosage level 1	2	3	Divisor	T_i	Variance	F
(b) Slope	— 1	0	1	40	$374 = T_b$	$3496.90 = B^2$	35.47
(c) Curvature	1	— 2	1	120	$-174 = T_c$	252.30	2.56
Totals ($U + S$)	455	729	829		Error $s_2^2 = 98.598$		
(a) S vs U	1	1	1	60	$-157 = T_a$	410.82	10.44
(ab) Parallel	— 1	0	1	40	$-18 = T_{ab}$	8.10	.21
(ac) Op. curv.	1	— 2	1	120	$-58 = T_{ac}$	28.03	.71
Totals ($U - S$)	— 53	— 33	— 71		Error $s_1^2 = 39.354$		

$y_u - y_s = -157 \times 2/60 = -5.2333$ (Eq. 31); $b = 374/.2219 \times 40 = 42.136$
$i = .2219$; $M' = -5.2333/42.136 = -.1242$ (Eq. 23b) (Eq. 7c)
$t_1 = t_2 = 2.052$ for $n = 27$, $P = .05$;
$s_2^2 t_2^2 = 415.20$; $B^2 - s_2^2 t_2^2 = 3081.70$
$C^2 = 3496.90/3081.70 = 1.1347$; $C = 1.0652$; $s_1^2 t_1^2 = 165.72$

$$X_L' = 1.1347(-.1242) \pm \frac{1.0652}{42.136}\sqrt{\frac{165.72 \times 2}{30} + \frac{415.20(-5.2333)^2}{3081.70}} \quad \text{(Eq. 108)}$$

$X_L' = -.1409 \pm .0970$; log(assumed potency in units per gram) $= 3.3010$
$M = 3.3010 - .1242 = 3.1768$; $X_L = 3.1601 \pm .0970$
Estimated potency = 1502 units per gram within limits of 1156 and 1808 units per gram

C. RELATIVE POTENCY FROM ALL-OR-NONE ASSAYS.

Graphic methods for estimating potency aid the experimenter in following the course of an investigation, in preliminary surveys of large bodies of data, and in many other cases. The short-cut described on pages 541-543 for the all-or-none response has been illustrated in Example 17a with the data of the antisterility assay for Vitamin E in Table LXVII and plotted in Figure 12, both being reprinted here from pages 149-150 of Volume II of *Vitamin Methods*.

The results of quantal assays may be considerably more irregular than in this example and still agree well with the hypothesis upon which the probit method is based. As the plotted values depart increasingly from exact linearity, the lines placed by inspection become less and less objective. The limitation this imposes upon the graphic technique has been examined empirically by Finney (79). Independent graphic estimates of the same data as fitted by 21 untrained subjects ranged from

TABLE LXVII. Antisterility Assay of Synthetic *dl,α*-Tocopherol ("Unknown" = U) in Terms of Synthetic *dl,α*-Tocopheryl Acetate ("Standard" = S), Where the Doses Represent the Tocopherol in Each Compound on a Molecular Weight Basis (Data of P. L. Harris)

Vitamin	Dose mg.	No. of rats	Fertile per cent	Log-dose (+1)	Response probits
S	0.4	8	25	.602	.433
	0.6	8	50	.778	5.00
	0.9	8	100	.954	
U	0.4	10	0	.602	
	0.6	12	16.7	.778	4.03
	0.9	10	60	.954	5.25

At point 5, $\tau = 0.75$ for Standard.
$x = 0.94$ for Unknown.
Log-potency $= -\overline{0.19}$ from graphic estimate
Antilog $(-0.19) = 0.65$ or in this assay *dl,α*-tocopherol was about 65% as potent as *dl,α*-tocopheryl acetate by graphic analysis

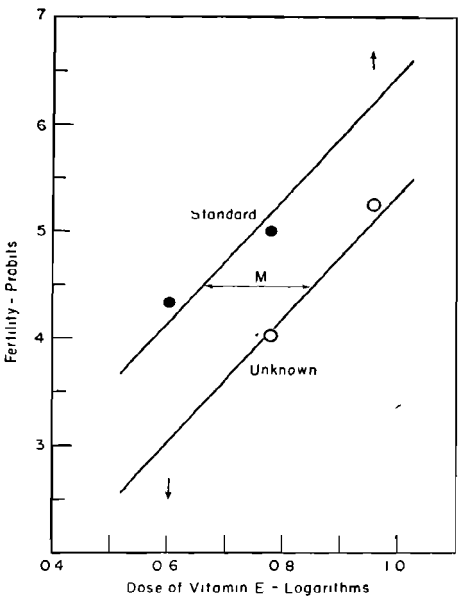

FIG. 12. Graphic evaluation of antisterility assay shown in Table LXVII.

a potency of 21.0 to 27.7 units per milligram. When the first computed estimate was determined from each pair of lines, the potency varied from 21.5 to 22.0 units per milligram. Thus only a single cycle of the maximum likelihood procedure produced a relatively stable, objective esti-

mate. The computation is further facilitated by tables which are available elsewhere (80).

Whether the observations agree with the fitted line within the sampling error is tested by χ^2, computed most conveniently by Eq. 63. The significance of χ^2 depends upon its degrees of freedom, which may be determined by the empirical rule on page 545. The discrepancy of each individual group is not evident when χ^2 is computed with Eq. 63. Especially if groups have had to be combined in estimating the degrees of freedom, groups with small expectations may contribute erratically to the estimate. This is avoided when χ^2 is determined directly from the number of observed and expected responses (43).

To compute χ^2 directly from the frequencies, the expected probit response Y at each observed log-dose X is determined with the computed curve and converted to the expected proportionate response P by a table of probits. P in turn is multiplied by N, the number of individuals in the group, to obtain NP, the expected number of positive reactions corresponding to each observed number D, where D/N is the observed proportionate response in the group. χ^2 is computed from the differences between the observed and expected frequencies of positive responses as

$$\chi^2 = S\left[\frac{(D - NP)^2}{NPQ}\right] \tag{109}$$

where $Q = 1 - P$. χ^2 is obtained as the sum of the k ratios enclosed in braces and has $k - 2$ degrees of freedom.

If either NP or $N - NP$ is less than 1 in a given group, it is added to the expected frequencies in the one or more groups at adjacent dosage levels until the expectation of both positive and negative responses is larger than 1. If groups must be combined at both low and high doses, larger expectations are preferred. The observed positive responses in these same groups are added similarly. When values have been pooled in this way, usually at one or both ends of the dosage-effect curve, D for the corresponding ratio in Eq. 109 is defined as the total of the observed positive responses, NP as the total of the expected positive responses, N as the total number in the combined group, and $Q = 1 - (NP/N)$.

In many all-or-none experiments, the individuals in each treated group are classified at the end of the test as either "dead" or "alive." Since death may be due to causes other than the test drug or toxicant, untreated control groups provide an essential check on the experiment. When there is "natural" mortality in the controls, not all the deaths in the treated groups can be attributed to the treatment. If the dosage-effect curve is to measure the effect of the drug or toxicant, the observed percentage dead at each dosage level must be corrected for the other

sources of mortality. These are usually assumed to act independently of the dose of drug, and the proportion killed by the treatment p is computed as

$$p = \frac{p' - p_c}{1 - p_c} \tag{110}$$

where p' is the observed proportion dead and p_c is the proportion dead in the controls. Probits based upon these corrected proportions are plotted against the log-dose for a graphic estimate of the dosage-mortality curve. In computing the approximate standard error of the log-LD50 with Eq. 53, N' is multiplied by $(1 - p_c)$. The calculation of the dosage-mortality curve when $p_c \neq 0$ has been described by Finney (43), to whose book the reader is referred.

Example 16c. The dosage-effect curve for vitamin E on page 539 agreed well with the observations when χ^2 was computed with Eq. 63. Although no other test is needed, the example may be used conveniently for illustrating the alternate calculation of χ^2 with Eq. 109. The expected probit for each log-dose was computed as $Y = -1.737 + 8.322X$, and each Y, in turn, converted to its expected P with a table of probits, as shown in the first three columns of Table LXVII. So few positive

TABLE LXVII. Alternate Calculation of χ^2 for the Dosage-Effect Curve in Table XLIII from the Expected and Observed Numbers of Deliveries

Log dose	Expected probit	Expected response	No. in group	No. positive			
				Observed	Expected	NPQ	$\frac{(D - PN)^2}{NPQ}$
x	Y	P	N	D	NP		
.574	3.040	.025	5	0	.12	1.67	.005
.699	4.080	.179	10	2	1.79		
.796	4.887	.455	10	4	4.55	2.48	.122
.875	5.545	.707	10	8	7.07	3.16	.099
1.000	6.585	.944	11	10	10.38		
1.176	8.050	.999	11	11	10.99		
							$\chi^2 = .226$

responses were expected at the smallest dose that the first two values were pooled, and so few negative responses at the two highest doses that the last three values were pooled. For the two lowest doses, $NP = 1.91$, $N = 15$ and $Q = 1 - 1.91/15 = .873$, giving $NPQ = 1.67$ and a contribution to χ^2 of $(2 - 1.91)^2/1.67 = .005$. The other ratios were determined similarly to obtain $\chi^2 = .226$ with one degree of freedom, which has the same probability (between 0.6 and 0.7) as the χ^2 with two degrees of freedom computed from the same data with Eq. 63.

REFERENCES

1. Fisher, R. A., and Yates, F., Statistical Tables for Biological, Agricultural and Medical Research. Third Edition. Oliver and Boyd, London, 1948.
2. Barlow's Tables of Squares, Cubes, Square Roots, Cube Roots and Reciprocals. Third Edition. Engineers Bookshop, New York, 1935.
3. Fisher, R. A., Statistical Methods for Research Workers. Tenth Edition. Hafner Publishing Co., New York, 1948.
4. Bliss, C. I., and Cattell, McK., *Ann. Rev. Physiol.* 5, 479 (1943).
5. Bliss, C. I., *J. Am. Pharm. Assoc.* 29, 465 (1940).
6. Irwin, J. O., *Suppl. J. Roy. Stat. Soc.* 4, 1 (1937).
7. Finney, D. J., *Suppl. J. Roy. Stat. Soc.* 9, 46 (1947).
8. Bliss, C. I., *Ann. N. Y. Acad. Sci.* 52, 877 (1950).
9. Jerne, N. K., and Wood, E. C., *Biometrics* 5, 273 (1949).
10. Burn, J. H., Finney, D. H., and Goodwin, L. G., Biological Standardization. Second Edition. Oxford University Press, London, 1950.
11. Coward, K. H., Biological Standardization of the Vitamins. Second Edition. Bailliere, Tindall and Cox, London, 1947.
12. Emmens, C. W., Principles of Biological Assay. Chapman and Hall, London, 1948.
13. Bacharach, A. L., Coates, M. E., and Middleton, T. R., *Biochem. J.* 36, 497 (1942).
14. Schønheyder, F., *Biochem. J.* 30, 890 (1936).
15. Leong, P. C., *Biochem. J.* 33, 1397 (1939).
16. Eisenhart, Churchill, *Biometrics* 3, 1 (1947).
17. Cochran, W. G., *Biometrics* 3, 22 (1947).
18. Bartlett, M. S., *Biometrics* 3, 39 (1947).
19. Fisher, R. A., The Design of Experiments. Fourth Edition. Hafner Publishing Co., New York, 1947.
20. Yates, F., *Empire J. Exptl. Agr.* 1, 129 (1933).
21. Gridgeman, N. T., *Biochem. J.* 37, 127 (1943).
22. Fieller, E. C., *Analyst* 72, 37 (1947).
23. Waddell, J., and Kennedy, G. H., *J. Assoc. Offic. Agr. Chemists* 30, 190 (1947).
24. U. S. Pharmacopoeia Vitamin Advisory Board. Collaborative Study of Vitamin D_3 proposed as the new international standard. Instructions to collaborators (Mimeographed) (1948).
25. Coward, K. H., and Kassner, E. W., *Biochem. J.* 35, 979 (1941).
26. Crampton, E. W., *J. Nutrition* 33, 491 (1947).
27. Irwin, J. O., *J. Hyg.* 43, 291 (1944).
28. Bliss, C. I., *Biometrics Bull.* 1, 57 (1945).
29. Bliss, C. I., *J. Am. Stat. Assoc.* 39, 479 (1944).
30. Grubbs, F. E., and Weaver, C. L., *J. Am. Stat. Assoc.* 42, 224 (1947).
31. Lord, E., *Biometrika* 34, 41 (1947).
32. Pearson, Karl, Tables for Statisticians and Biometricians. Part II. Biometric Laboratory, University College, London, 1931.
33. Bliss, C. I., *Biometrics* 3, 69 (1947).
34. Gridgeman, N. T., The Estimation of Vitamin A. Lever Bros. and Unilever, Ltd., 1944.
35. Smith, K. W., Marks, H. P., Fieller, E. C., and Broom, W. A., *Quart. J. Pharm. Pharmacol.* 17, 108 (1944).

36. Gridgeman, N. T., *Quart. J. Pharm. Pharmacol.* **18**, 15 (1945).
37. Bliss, C. I., *J. Am. Stat. Assoc.* **35**, 498 (1940).
38. Finney, D. J., *J. Hyg.* **45**, 397 (1947).
39. Bliss, C. I., and Marks, H. P., *Quart. J. Pharm. Pharmacol.* **12**, 82; **12**, 182 (1939).
40. Gaddum, J. H., *Med. Research Council (Brit.) Special Rept. Ser.* **No. 183** (1933).
41. Bliss, C. I., *Quart. J. Pharm. Pharmacol.* **11**, 192 (1938).
42. Miller, L. C., Bliss, C. I., and Braun, H. A., *J. Am. Pharm. Assoc.* **28**, 644 (1939).
43. Finney, D. J., Probit Analysis. 2nd Ed. Cambridge University Press, 1952.
44. Litchfield, J. T., and Fertig, J. W., *Bull. Johns Hopkins Hosp.* **69**, 276 (1941).
45. Mason, K. E., *J. Nutrition* **23**, 59 (1942).
46. deBeer, E. J., *J. Pharmacol. Exptl. Therap.* **85**, 1 (1945).
47. Litchfield, J. T., Jr., and Wilcoxon, F., *J. Pharmacol. Exptl. Therap.* **96**, 99 (1949).
48. Pugsley, L. I., Wills, G., and Crandall, W. A., *J. Nutrition* **28**, 365 (1944).
49. Bartlett, M. S., *Suppl. J. Roy. Stat. Soc.* **4**, 137 (1937).
50. Harris, L. J., *League Nations Bull. Health Organisation* **9**, 402 (1940-41).
51. Wood, E. C., *Nature* **155**, 632 (1945).
52. U. S. Pharmacopoeia XIV, pp. 737, 750. Mack Publishing Co., Easton, Pa., 1950.
53. Snell, E. E., *Physiol. Revs.* **28**, 255 (1948).
54. Doede, D. R., *Yale J. Biol. and Med.* **17**, 595 (1945).
55. Emery, W. B., McLeod, N., and Robinson, F. A., *Biochem. J.* **40**, 426 (1946).
56. Wood, E. C., *Analyst* **71**, 1 (1946).
57. Finney, D. J., *Quart. J. Pharm. Pharmacol.* **18**, 77 (1945).
58. Wood, E. C., and Finney, D. J., *Quart. J. Pharm. Pharmacol.* **19**, 112 (1946).
59. Bliss, C. I., *Ann. Math. Stat.* **17**, 232 (1946).
60. Cochran, W. G., *Suppl. J. Roy. Stat. Soc.* **4**, 102 (1937).
61. Harris, P. L., Jensen, J. L., Joffe, M., and Mason, K. E., *J. Biol. Chem.* **156**, 491 (1944).
62. Waddell, J., and Kennedy, G. H., *J. Assoc. Offic. Agr. Chemists* **30**, 190 (1947).
63. Bliss, C. I., *Poultry Sci.* **24**, 534 (1945).
64. American Standards Association, N. Y., American War Standard Z1.3, 1942.
65. Dudding, B. P., and Jennett, W. J., British Standard 600R: 1942. British Standards Inst., London.
66. American Society for Testing Materials, Philadelphia, Am. Soc. Testing Materials Manual on Presentation of Data, Suppl. B, p. 47, 1945.
67. Merrington, M., and Thompson, C. M., *Biometrika* **33**, 73 (1943).
68. Wald, A., and Wolfowitz, J., *Ann. Math. Stat.* **17**, 208 (1946).
69. Fisher, R. A., *Biometrics* **3**, 65 (1947).
70. Bliss, C. I., *Ann. Applied Biol.* **24**, 815 (1937).
71. Bliss, C. I., Result of the First U.S.P. Collaborative Plate-Assay of Penicillin. Mimeographed by U.S.P. Committee of Revision, 1945.
72. Knudsen, L. F., and Tolle, C. D., *J. Assoc. Offic. Agr. Chemists* **33**, 665 (1940).
73. Bliss, C. I., *J. Assoc. Offic. Agr. Chemists* **29**, 396 (1946).
74. Bliss, C. I., *J. Am. Pharm. Assoc., Sci. Ed.* **33**, 225 (1944).
75. Miller, L. C., *J. Am. Pharm. Assoc., Sci. Ed.* **33**, 245 (1944).
76. Finney, D. J., *J. General Microbiol.* **5**, 223 (1951).

77. Finney, D. J., *Acta Pharmacol. Toxicol.* **8**, 55 (1952).
78. Fieller, E. C., *Quart. J. Pharm. Pharmacol.* **17**, 117 (1944).
79. Finney, D. J., *J. Pharmacol. Exptl. Therap.* **104**, 440 (1952).
80. Bliss C. I. and Calhoun D. W. An Outline of Biometry. Yale Co-Operative Corporation, New Haven, Conn., 1953.

Index to Equations

Index to Work-Forms and Reference Tables

Subject Index

D

E