ρ is estimated, as in the graphical method, either from the ratio of slopes or from the horizontal distance between parallel lines.

The three forms of estimation may be summarized as follows:

(i) Standard curve: Regression equation, not necessarily linear, assumed to remain fixed in position;

(ii) Standard slope: Regression equation on log dose assumed linear, with constant regression coefficient;

(iii) Simultaneous trial, two doses of each preparation: Regression on dose metameter assumed linear, but each assay provides its own estimate of all parameters;

(iv) Simultaneous trial, three or more doses of each preparation: As (iii), but each assay also gives a test of deviations from linearity.

The standard curve and standard slope methods will not be discussed further.

4

Parallel line assays

4.1 Unsymmetric designs

The most widely used type of simultaneous trial assay is that for which a simple response metameter has a homoscedastic linear regression on log dose. For such an assay, the condition of similarity requires the lines for the standard and test preparations to be parallel (§3.7). As will become apparent in Chapters 5 and 6, symmetry in the number and spacing of doses and in the allocation of subjects to doses usually improves precision and eases computation. Nevertheless, an accident may convert a symmetric design into an unsymmetric, or shortage of material may force adoption of an unsymmetric design. A general unsymmetric assay, such as is discussed below, is also the best illustration of the whole structure of the computations.

4.2 Data for an unsymmetric assay

Table 4.2.1 (British Standards Institution, 1940) relates to an assay of vitamin D_3 in cod-liver oil by means of its antirachitic activity in chickens, using percentage bone ash as the response. When the measured responses are percentages, both non-linearity and heteroscedasticity of the regression are likely, at least at extreme doses (§3.9). In this assay, almost all the responses lay between 30 percent and 45 percent, and no difficulties of statistical invalidity (§§4.5, 4.6) were encountered in an analysis based on a linear regression of response on log dose. Although no response metameter was needed, a linear transformation (commonly termed a *coding*) was applied. Each bone ash percentage, u, was transformed by

$$y = 10(u - 30), \tag{4.2.1}$$

so that a response 33·5 was coded as 35, etc. Such coding can reduce the magnitudes of quantities used in calculations, remove decimal digits, and make most values positive, all of which are conveniences in desk calculation though seldom worth while on a computer. Responses recorded to halves or quarters of arbitrary units can be coded by multiplication by 2 or 4. A linear metametric transformation does not affect scedasticity, linearity, tests of validity, or potency estimates. Table 4.2.1 contains the coded data for the vitamin D_3 assay; by inversion of equation (4.2.1), the original percentages are recoverable:

$$u = 30 + 0{\cdot}1y. \tag{4.2.2}$$

One good feature of the assay is that the doses were at equal logarithmic spacing: for both preparations, the ratio of successive doses was 5/3. With a logarithmic dose metameter, such a choice can ensure that the dose range is adequately

TABLE 4.2.1 Responses in an assay of cod-liver oil for vitamin D_3

	Dose of standard preparation, S (BSI units per 100 g food)			Dose of test preparation, T (mg oil per 100 g food)			
	5·76	9·6	16	32·4	54	90	150
	35	62	116	20	26	57	140
	30	67	105	39	60	89	133
	24	95	91	16	48	103	142
	37	62	94	27	−8	129	118
	28	54	130	−12	46	139	137
	73	56	79	2	77	128	84
	31	48	120	31		89	101
	21	70	124			86	
	−5	94					
		42					
n	9	10	8	7	6	8	7
S y	274	650	859	123	249	820	855
$\bar{y}$	30·4	65·0	107·4	17·6	41·5	102·5	122·1
x	−2	0	2	−3	−1	1	3

covered and simplify the arithmetic. The calculations could be executed with $\log_{10} z$ as the metameter. If instead

$$x = \log_\epsilon z - \log_\epsilon 9{\cdot}6 \tag{4.2.3}$$

for S, and

$$x = \log_\epsilon z - \tfrac{1}{2}(\log_\epsilon 54 + \log_\epsilon 90) \tag{4.2.4}$$

for T, where

$$\epsilon = (5/3)^{1/2}, \tag{4.2.5}$$

the doses are represented by the simple integers shown in the last line of Table 4.2.1. Alternatively, in order to avoid negative values, metameters might have been chosen so that the values of x were 0, 1, 2, for S and 0, 1, 2, 3 for T; equations (4.2.3), (4.2.4) have the advantage of making $\bar{x}_S$, $\bar{x}_T$ almost zero (not exactly, because of unequal numbers of subjects per dose). The essential feature of the metameter scale is that the same base of logarithms is used for both preparations. The choice of base and the addition of different quantities to the logarithms for the two preparations affect only the intermediate arithmetic; a simple final adjustment removes their effects.

4.3 The dose–response diagram

A diagram (Fig. 4.3.1) showing mean responses plotted against x leads to a rapid estimation of potency; it also protects against gross errors or misinterpretations of the statistical analysis. The experienced user of assay techniques may often dispense with the diagram, at least for symmetric designs, because he can visualize its form without drawing it, but to others a sketch is practically essential.

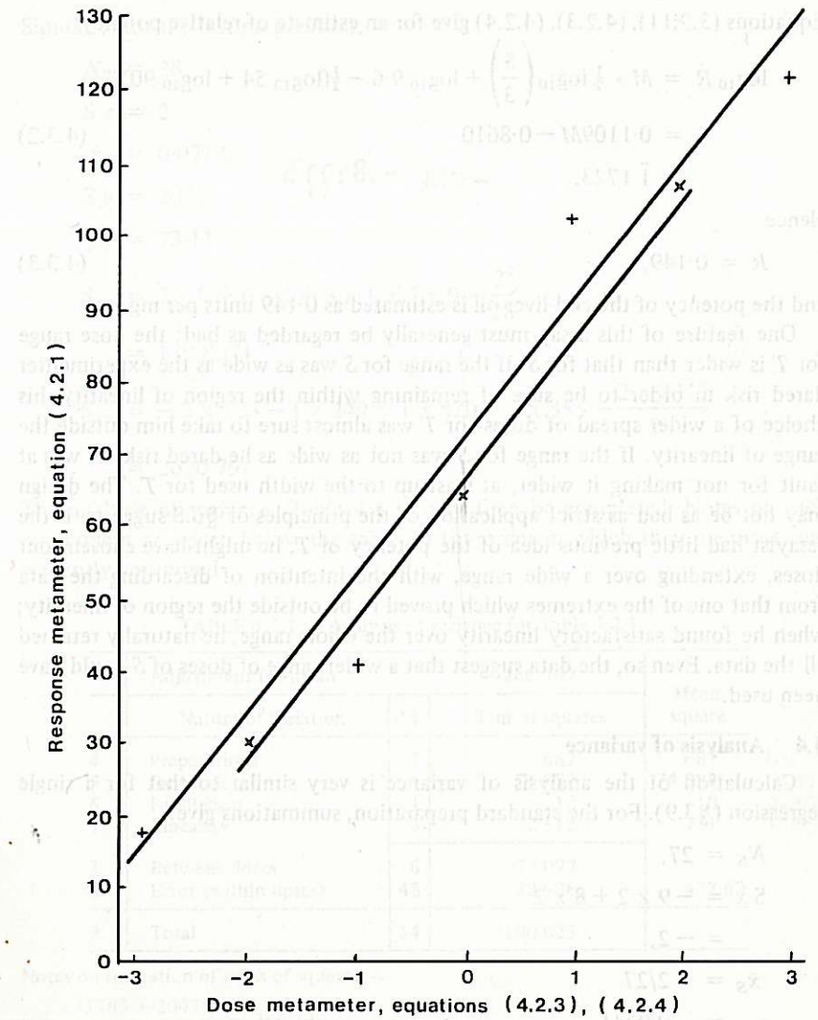

Fig. 4.3.1 Linear dose–response regressions for the assay of vitamin D_3, Table 4.2.1
×: Mean responses to standard preparation
+: Mean responses to test preparation
The straight lines are those drawn by eye (§4.3), but the calculated equations (§4.11) are almost identical with them.

In Fig. 4.3.1, two parallel lines have been drawn by eye so as to fit the points approximately. The horizontal distance between these lines roughly estimates the difference in x between doses giving equal responses:

$$M = 0{\cdot}3. \tag{4.3.1}$$

Equations (3.9.11), (4.2.3), (4.2.4) give for an estimate of relative potency:

$$\log_{10} R = M \times \tfrac{1}{2}\log_{10}\left(\frac{5}{3}\right) + \log_{10} 9{\cdot}6 - \tfrac{1}{2}(\log_{10} 54 + \log_{10} 90)$$
$$= 0{\cdot}1109M - 0{\cdot}8610 \qquad (4.3.2)$$
$$= \bar{1}{\cdot}1723.$$

Hence

$$R = 0{\cdot}149, \qquad (4.3.3)$$

and the potency of the cod-liver oil is estimated as 0·149 units per mg.

One feature of this assay must generally be regarded as bad: the dose range for T is wider than that for S. If the range for S was as wide as the experimenter dared risk in order to be sure of remaining within the region of linearity, his choice of a wider spread of doses for T was almost sure to take him outside the range of linearity. If the range for S was not as wide as he dared risk, he was at fault for not making it wider, at least up to the width used for T. The design may not be as bad as strict application of the principles of §6.8 suggests. If the assayist had little previous idea of the potency of T, he might have chosen four doses, extending over a wide range, with the intention of discarding the data from that one of the extremes which proved to be outside the region of linearity; when he found satisfactory linearity over the whole range, he naturally retained all the data. Even so, the data suggest that a wider range of doses of S could have been used.

4.4 Analysis of variance

Calculation of the analysis of variance is very similar to that for a single regression (§3.9). For the standard preparation, summations give:

$$N_S = 27,$$
$$\mathrm{S}x = -9 \times 2 + 8 \times 2$$
$$= -2,$$
$$\bar{x}_S = -2/27$$
$$= -0{\cdot}0741,$$
$$\mathrm{S}y = 1783,$$
$$\bar{y}_S = 66{\cdot}04,$$
$$\mathrm{S}_{xx} = 9 \times 4 + 8 \times 4 - \frac{(-2)^2}{27}$$
$$= 67{\cdot}8519,$$
$$\mathrm{S}_{xy} = -2 \times 274 + 2 \times 859 - \frac{(-2) \times 1783}{27}$$
$$= 1302{\cdot}07.$$

Similarly, for the test preparation,

$$N_T = 28,$$
$$\mathrm{S}x = 2,$$
$$\bar{x}_T = 0{\cdot}0714,$$
$$\mathrm{S}y = 2047,$$
$$\bar{y}_T = 73{\cdot}11,$$
$$\mathrm{S}_{xx} = 7 \times 9 + 6 \times 1 + 8 \times 1 + 7 \times 9 - \frac{2^2}{28}$$
$$= 139{\cdot}8571,$$
$$\mathrm{S}_{xy} = -3 \times 123 - 1 \times 249 + 1 \times 820 + 3 \times 855 - \frac{2 \times 2047}{28}$$
$$= 2620{\cdot}79.$$

The analysis of variance, Table 4.4.1, may then be completed. Notes on each component are given below the table, in the order in which they are most conveniently computed.

TABLE 4.4.1 Analysis of variance for Table 4.2.1

	Nature of variation	d.f.	Sum of squares	Mean square
1	Adjustment for mean		266 707	
4	Preparations	1	687	687
5	Regression	1	74 088	74 088
6	Parallelism	1	10	10
7	Linearity	3	2 312	771
3	Between doses	6	77 097	
8	Error (within doses)	48	22 928	477·67
2	Total	54	100 025	

Notes on formation of sums of squares:–

(1) $\dfrac{(1783 + 2047)^2}{55} = 266\,707$

(2) $35^2 + 30^2 + 24^2 + \ldots + 84^2 + 101^2 - 266\,707 = 100\,025$

(3) $\dfrac{274^2}{9} + \dfrac{650^2}{10} + \dfrac{859^2}{8} + \dfrac{123^2}{7} + \dfrac{249^2}{6} + \dfrac{820^2}{8} + \dfrac{855^2}{7} - 266\,707 = 77\,097$

(4) $\dfrac{1783^2}{27} + \dfrac{2047^2}{28} - 266\,707 = 687$

(5) Pooled regression component, given by

$$\frac{(\Sigma \mathrm{S}_{xy})^2}{\Sigma \mathrm{S}_{xx}} = \frac{(1302{\cdot}07 + 2620{\cdot}79)^2}{67{\cdot}8519 + 139{\cdot}8571}$$
$$= \frac{(3922{\cdot}86)^2}{207{\cdot}7090}$$
$$= 74\,088$$

(6) Difference between fitting two independent regression coefficients and one pooled value is

$$\Sigma\left\{\frac{(S_{xy})^2}{S_{xx}}\right\} - \frac{(\Sigma S_{xy})^2}{\Sigma S_{xx}} = \frac{1302{\cdot}07^2}{67{\cdot}8519} + \frac{2620{\cdot}79^2}{139{\cdot}8571} - 74\,088 = 10$$

(7) $77\,097 - (687 + 74\,088 + 10) = 2312$

(8) $100\,025 - 77\,097 = 22\,928.$

The components of the analysis must next be examined. Unless evidence of its unsuitability is found, the mean square from the error line,

$$s^2 = 477{\cdot}67, \qquad (4.4.1)$$

will be used as the basic variance estimate.

4.5 Scedasticity

The error sum of squares comprises contributions from each of the seven doses. These may be examined separately in a study of evidence for heteroscedasticity. Tests of significance for other validity tests (§§4.6–4.8), and assessment of precision of the potency estimate (§4.14), in theory require that $\sigma^2(y)$ be the same for all doses (§3.10), though experience shows that even quite large departures from homoscedasticity do not matter much. In more complex designs, to test the homogeneity of variance may be impracticable, but care in the preliminary investigation and in the choice of doses will generally remove the risk of heteroscedasticity so severe as to disturb the estimation seriously.

TABLE 4.5.1 Test of variance heterogeneity for Table 4.2.1

	f_i	Sum of squares	s_i^2	$\log s_i^2$
S	8	3268	408·5	2·611
	9	2808	312·0	2·494
	7	2280	325·7	2·513
T	6	1854	309·0	2·490
	5	4356	871·2	2·940
	7	5392	770·3	2·887
	6	2971	495·2	2·695
	48	22929	477·7	2·679

Table 4.5.1 shows the seven mean squares, s_i^2, with their degrees of freedom, f_i; the last line of the table contains the total number of degrees of freedom, f, and the pooled mean square, s^2. A final column shows the logarithms of the mean squares. Bartlett's test (§3.11) then gives, by equations (3.11.4) and (3.11.1)

$$C = 1 + \left(\frac{1}{5} + \frac{2}{6} + \frac{2}{7} + \frac{1}{8} + \frac{1}{9} - \frac{1}{48}\right) \Big/ 18 = 1{\cdot}0575$$

$$\chi^2_{[6]} = 2{\cdot}3026 \times 1{\cdot}648/1{\cdot}0575 = 3{\cdot}59;$$

Tabel IV

χ^2 is so well below the 0·05 significance level (12·6) that adjustment by the factor C could have been omitted. The data cause no worry about heteroscedasticity.

4.6 Linearity

A preliminary investigation will be presumed to have established that, over a range of responses such as occurs here, the regression of bone ash percentage on log dose is practically linear. The mean square for 'Linearity', or, more fully, 'Deviations from linearity' should still be examined, as a check that nothing has seriously disturbed this linearity. Unless accompanied by other danger signals, a mean square that is large relative to the error would most probably indicate *statistical invalidity*, that is to say inappropriateness of the form of analysis adopted. For example, a bad choice of doses for either preparation might take most of the observations off the linear portion of the response curve; the conditions of similarity and monotony might be fulfilled, so that in theory the data would still be suitable for an assay, but the assumption of a linear regression would no longer be justified. This need not be evidence against the inherent comparability of the two preparations. The assay might be rejected, however, because a satisfactory linearizing transformation could not be found without more extensive data; it might be rejected because changes in dose had so little effect on response as to make any estimate of ρ hopelessly imprecise; or it might still be usable (see also §4.22).

The mean square for linearity in Table 4.4.1 is greater than that for error, but not significantly so; the ratio is not great enough to occasion any alarm.

4.7 The difference in preparations

In an ordinary experiment for comparing treatments, major interest attaches to differences between treatment means. Here the difference between S and T in their mean responses is not of intrinsic interest. A large difference, however, will seldom arise unless the responses to either the lowest or the highest doses of T lie far outside the range of responses to S, though the converse is not necessarily true. As already implied (§4.3), this should not happen: if it does, either the range of doses for S ought to have been wider, or that for T was too wide and extended beyond the region of linearity. Moreover, as will be apparent from §4.14, a large difference in mean responses will decrease the precision of potency estimation (§6.8).

In this assay, at both extremes of dose the responses to T lie outside the range for S, and in the mean response these extremes compensate for one another. Though the mean square for preparations is only a little greater than the error mean square, the assay is certainly open to criticism on account of the choice of doses of T. A large mean square for the difference in preparations is always a danger signal, but a small one is no assurance that all is well. Results of an assay like the present should be treated with some reserve, and the response diagram

should be inspected for any indications of non-linearity at the extremes of the test preparation regression.

4.8 Parallelism

If other tests had disclosed no significantly large mean squares for linearity or preparations, a large mean square for the component based on deviations from parallelism would indicate *fundamental invalidity* of the assay. When log dose is used as a metameter, an essential condition for an analytic dilution assay is that the regression curves are parallel (§3.3). If these curves are linear (§3.7), non-parallelism would violate the condition of similarity. Whether the initial assumption that T behaved as a dilution of S was inherently false or whether it had been obscured by an impurity in one preparation, the data would have to be discarded. This is without prejudice to the possibility of deliberately using non-parallel regressions for other types of assay, as suggested by Thompson (1948). Table 4.4.1 shows no evidence of deviations from parallelism.

If danger signals appear simultaneously in several validity tests, assignment of the cause to one explanation may be impossible. The whole assay is then suspect, and should be discarded; whether it is fundamentally invalid or statistically intractable matters little, except in so far as the planning and experimental technique for the next assay may be affected.

4.9 Regression

No assay should be undertaken without strong prior belief in the existence of a regression, without which the dose–response relation is useless for estimating potency. In a good assay, therefore, the variance ratio for the regression component will generally be highly significant. Here it is

$$F = \frac{74\,088}{477{\cdot}67} = 155{\cdot}1.$$

Only when F is large are fiducial limits to the potency narrow enough to be useful (§4.14).

4.10 Significance levels

In the validity tests described in §§4.5–4.8, though a test of statistical significance at a probability of 0·05 was implied, each conclusion was so clear that any reasonable probability would have given the same answer. The assayist need not use the same probability here as for the fiducial limits. What level is ideal?

Some might think that stringent tests should be applied, especially for parallelism, because of the importance of rejecting invalid analyses: perhaps a probability of 0·10 should be used instead of 0·05. Others might think this extravagant, because many good sets of data would be rejected on account of the mischances of random sampling. Experience suggests that, in a well planned assay, even fairly large deviations from the strict theoretical requirements of

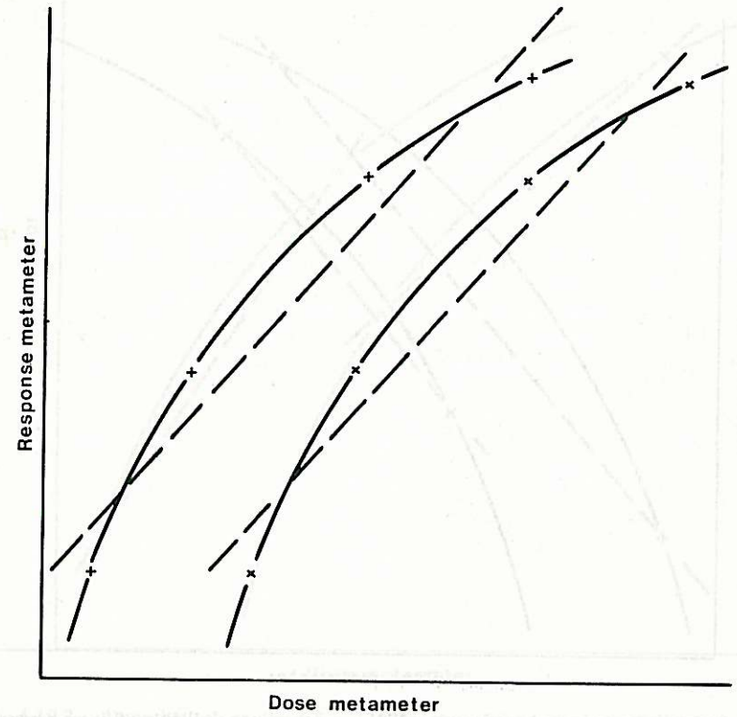

Fig. 4.10.1 A successful choice of doses permits a valid estimation of potency from parallel linear regressions, even though the true regressions are curved
×: Mean responses to standard preparation
+: Mean responses to test preparation
Full lines indicate regression curves, broken lines are hypothetical estimated linear regressions.

statistical validity will have little effect on the estimate of ρ and not much on the assessment of its precision. If the design is symmetric, with the same number of doses for both preparations and equal numbers of subjects at every dose (Chapter 5), and if the assayist guesses his doses of T so successfully that they are almost exactly equivalent to those of S, a determination of R from the horizontal distance between linear regressions will be valid, even though the true regression is curved (Fig. 4.10.1). In such an assay, the assessment of error may be seriously upset (but see §5.11). With the same true regression, a bad choice of doses may give apparent parallelism but a hopelessly biased estimate (Fig. 4.10.2), or complete non-parallelism (Fig. 4.10.3), in spite of the fundamental validity of the assay.

Perhaps the chief danger is that of fundamental invalidity, and a reasonably stringent parallelism test therefore seems desirable. A deviation from parallelism significant at a probability of 0·05 should be regarded as sufficient cause for

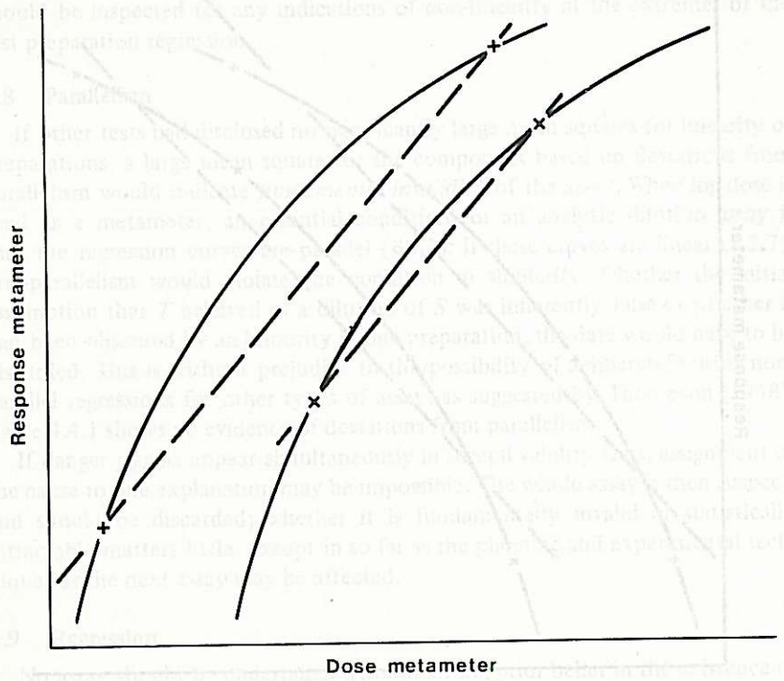

Fig. 4.10.2 Hypothetical results of a (2, 2) assay showing parallelism of linear regressions but biased estimation of potency; regression curves conform to the condition of similarity but dose intervals are unequal

×: Mean responses to standard preparation
+: Mean responses to test preparation

Full lines indicate regression curves, broken lines are hypothetical estimated linear regressions.

rejection of an assay, unless extenuating circumstances not only explain the situation but ensure a statistically valid analysis. Usually the history of an assay technique for estimating the potency of test preparations with respect to a particular effective constituent, by an accepted experimental procedure and on a known stock of subjects, provides a strong presumption of parallelism; without this, even a deviation significant at a probability of 0·10 should be a little suspect.

When experience of a technique has given grounds for belief in similarity and in the statistical validity of the method of evaluation of the data, and also when the current assay gives no evidence against parallelism, less stringent tests for heteroscedasticity, deviations from linearity, and the difference between preparations might be allowed. As a working rule to be applied and interpreted intelligently, not uncritically, acceptance of the analysis as statistically valid unless one or more of these criteria are significant at a probability of 0·01 seems reasonable. The question is not entirely one of statistics; the knowledge of the chemist and

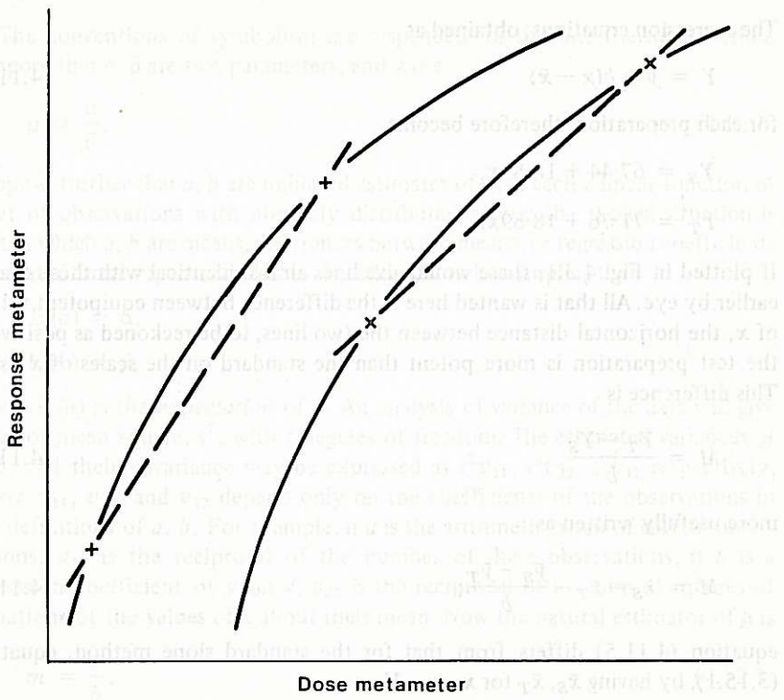

Fig. 4.10.3 Hypothetical results of a (2, 2) assay showing non-parallelism of linear regressions; regression curves conform to the condition of similarity but doses fail to correspond

×: Mean responses to standard preparation
+: Mean responses to test preparation

Full lines indicate regression curves, broken lines are hypothetical estimated linear regressions.

biologist about the materials of the assay must also be taken into account. For routine assays, quality control techniques may help interpretation of the criteria of statistical validity (§5.10).

4.11 Potency estimation

The cod-liver oil assay is free from evidence of invalidity, and estimation can proceed. The regression coefficient is estimated as

$$b = \frac{\Sigma S_{xy}}{\Sigma S_{xx}} \qquad (4.11.1)$$

$$= \frac{3922{\cdot}86}{207{\cdot}7090}$$

$$= 18{\cdot}89. \qquad (4.11.2)$$

The regression equations, obtained as

$$Y = \bar{y} + b(x - \bar{x}) \tag{4.11.3}$$

for each preparation, therefore become

$$Y_S = 67{\cdot}44 + 18{\cdot}89x,$$

$$Y_T = 71{\cdot}76 + 18{\cdot}89x.$$

If plotted in Fig. 4.3.1, these would give lines almost identical with those drawn earlier by eye. All that is wanted here is the difference between equipotent values of x, the horizontal distance between the two lines, to be reckoned as positive if the test preparation is more potent than the standard on the scales of x used. This difference is

$$M = \frac{Y_T - Y_S}{b}, \tag{4.11.4}$$

more usefully written as

$$M = \bar{x}_S - \bar{x}_T - \frac{\bar{y}_S - \bar{y}_T}{b}; \tag{4.11.5}$$

equation (4.11.5) differs from that for the standard slope method, equation (3.15.1), by having $\bar{x}_S, \bar{x}_T$ for x_S, x_T. Here

$$\begin{aligned} M &= -0{\cdot}0741 - 0{\cdot}0714 - \frac{66{\cdot}04 - 73{\cdot}11}{18{\cdot}89} \\ &= -0{\cdot}1455 + \frac{7{\cdot}07}{18{\cdot}89} \\ &= 0{\cdot}2288. \end{aligned} \tag{4.11.6}$$

This must be transformed by equation (4.3.2) to give

$$\log_{10} R = \bar{1}{\cdot}1644,$$

whence

$$R = 0{\cdot}1460,$$

as compared with the graphical estimate, 0·149, in equation (4.3.3).

4.12 Fieller's theorem

In equation (4.11.5), $(\bar{x}_S - \bar{x}_T)$ is a constant set by the choice of doses and numbers of subjects. Consequently, fiducial limits for M depend upon the second term, $(\bar{y}_S - \bar{y}_T)/b$. They are obtained from an important theorem first fully enunciated by Fieller (1940), though others (Bliss, 1935b) had earlier stated the particular case for zero covariance.

The conventions of symbolism are suspended for this and the next section. Suppose that α, β are two parameters, and write

$$\mu = \frac{\alpha}{\beta}. \tag{4.12.1}$$

Suppose further that a, b are unbiased estimates of α, β, each a linear function of a set of observations with normally distributed errors; the typical situation is that in which a, b are means, differences between means, or regression coefficients calculated from experimental data. Freedom from bias implies that

$$\mathrm{E}(a) = \alpha,$$

$$\mathrm{E}(b) = \beta,$$

where E (a) is the *expectation* of a. An analysis of variance of the data will give an error mean square, s^2, with f degrees of freedom. The estimated variances of a, b, and their covariance may be expressed as s^2v_{11}, s^2v_{22}, s^2v_{12} respectively, where v_{11}, v_{22}, and v_{12} depend only on the coefficients of the observations in the definitions of a, b. For example, if a is the arithmetic mean of certain observations, v_{11} is the reciprocal of the number of these observations; if b is a regression coefficient of y on x, v_{22} is the reciprocal of the sum of squares of deviations of the values of x about their mean. Now the natural estimator of μ is

$$m = \frac{a}{b}. \tag{4.12.2}$$

Fieller's theorem states that upper and lower fiducial limits to μ are(*)

$$m_L, m_U = \left[m - \frac{gv_{12}}{v_{22}} \pm \frac{ts}{b} \left\{ v_{11} - 2mv_{12} + m^2v_{22} - g\left(v_{11} - \frac{v_{12}^2}{v_{22}}\right) \right\}^{1/2} \right] \Big/ (1-g), \tag{4.12.3}$$

where

$$g = \frac{t^2s^2v_{22}}{b^2} \tag{4.12.4}$$

and t is the t-deviate with f degrees of freedom (Appendix Table I). The proof follows from consideration of the expression $(a - \mu b)$. For any μ, this also is a linear function of the observations; it has expectation

$$\mathrm{E}(a - \mu b) = \alpha - \mu\beta = 0, \tag{4.12.5}$$

and an estimated variance

$$\mathrm{Var}(a - \mu b) = s^2(v_{11} - 2\mu v_{12} + \mu^2 v_{22}), \tag{4.12.6}$$

(*)The symbol '±' generally introduces a standard error. In Fieller's theorem or one of its analogues, the symbol indicates the alternative operations of subtraction and addition needed for the lower and upper limits.

with f degrees of freedom. Hence (assuming normality), with probability appropriate to the t-deviate.

$$(a-\mu b)^2 \leqslant t^2 s^2 (v_{11} - 2\mu v_{12} + \mu^2 v_{22}); \tag{4.12.7}$$

the equality sign gives a quadratic equation in μ, whose solution is (4.12.3).

When b is large relative to its standard error, g will be small; if g can be neglected, equation (4.12.3) becomes

$$m_L, m_U = m \pm ts(v_{11} - 2mv_{12} + m^2 v_{22})^{1/2}/b, \tag{4.12.8}$$

a formula equivalent to using the expression

$$\mathrm{Var}(m) = s^2(v_{11} - 2mv_{12} + m^2 v_{22})/b^2 \tag{4.12.9}$$

as though it were the variance of m. This variance formula, obtained in other ways, is often used as a way of attaching a standard error to a ratio, especially in the form applicable when $v_{12} = 0$:

$$\frac{\mathrm{Var}(m)}{m^2} = s^2\left(\frac{v_{11}}{a^2} + \frac{v_{22}}{b^2}\right). \tag{4.12.10}$$

The approximation (4.12.8) is adequate if g is less than 0·05, which for limits at probability 0·95 requires b to be at least nine times its standard error. For large g, equation (4.12.3) is essential: the approximation much underestimates the width of the fiducial interval when g exceeds 0·2. New complications arise if g exceeds 1·0, as b then does not differ significantly from zero (Fieller, 1954). Detailed study of the inequality (4.12.7) shows that when $g = 1{\cdot}0$ one of m_L, m_U becomes infinite. When $g > 1{\cdot}0$, the range of values of μ that satisfy the inequality remains infinite, and the limits set by (4.12.3) become exclusive instead of inclusive: the assertion made with the chosen probability is that μ lies outside the interval m_L, m_U. The logic is sound, but the practical value of an assay with $g \geqslant 1{\cdot}0$ is small.

Biological assay often requires that fiducial limits be assigned to a ratio of two means, a ratio of two regression coefficients, or a horizontal distance between two parallel regression lines. Although many publications on bioassay have used equation (4.12.9) or some equivalent formula, the extra labour of calculating from (4.12.3) is so small that routine employment of the full formula is preferable to argument about the adequacy of the approximation. Even for desk calculation, this should be standard practice; when a computer is programmed for assay analyses, there is no excuse for not incorporating equation (4.12.3).

4.13 Analogues of Fieller's theorem

Valuable though Fieller's theorem is, it is not applicable to every bioassay in which estimation involves a ratio. Complications arise when more than one error mean square must be used; this situation will not be encountered until Chapter 9, but generalized theorems based on the Behrens distribution are for convenience presented here. First the nature of this distribution itself, a generalization of the t distribution, must be considered.

Again suppose that a, b are unbiased estimates of α, β, defined as linear functions of observations with normally distributed errors, but now with estimated variances

$$\left.\begin{aligned} \mathrm{Var}(a) &= s_1^2 v_{11}, \\ \mathrm{Var}(b) &= s_2^2 v_{22}, \end{aligned}\right\} \tag{4.13.1}$$

and zero covariance, where s_1^2, s_2^2 are independent mean squares with f_1, f_2 degrees of freedom respectively. The variance of $(a-b)$ is estimated as

$$\mathrm{Var}(a-b) = s_1^2 v_{11} + s_2^2 v_{22}. \tag{4.13.2}$$

If s_1^2, s_2^2 were the same mean square, the deviation of $(a-b)$ from its expectation divided by the estimated standard error,

$$\frac{(a-b)-(\alpha-\beta)}{(s_1^2 v_{11} + s_2^2 v_{22})^{1/2}}, \tag{4.13.3}$$

would follow the t distribution with f degrees of freedom. When s_1^2, s_2^2 are independent mean squares, however, this is no longer true, and the distribution of the ratio (4.13.3) is the Behrens distribution (Appendix Table III; Fisher and Yates, 1963, Table V_1). The ratio, the Sukhatme d-statistic, has a distribution defined in terms of the degrees of freedom f_1, f_2, and the angle θ such that

$$\tan^2\theta = \frac{s_1^2 v_{11}}{s_2^2 v_{22}}. \tag{4.13.4}$$

When θ is 0°, d is distributed as t with f_2 degrees of freedom; when θ is 90°, it is distributed as t with f_1 degrees of freedom. For other angles, the value of d for any probability is generally (but not always) intermediate between $t_{[f_1]}$ and $t_{[f_2]}$. When $f_1 = f_2$, the value of d for any probability is about equal to, but a little less than, the corresponding t, irrespective of the size of θ. The d-test is appropriate for testing a difference between two means or two regression coefficients whose variances are based on independent mean squares that cannot be assumed estimates of the same population variance and therefore must not be pooled. To attempt to refer the ratio (4.13.3) to the t distribution may mislead seriously if f_1 and f_2 are small.

Suppose now that a and b are known to be estimates of the same quantity ($\alpha = \beta$). They might be two estimates of the same mean, or two estimates of a regression coefficient, from different sets of observations with different variances. The most precise estimate that can be compounded of a and b is a weighted mean having the reciprocals of the variances as weights; this is $\bar{a}$, where

$$\bar{a}\left(\frac{1}{s_1^2 v_{11}} + \frac{1}{s_2^2 v_{22}}\right) = \frac{a}{s_1^2 v_{11}} + \frac{b}{s_2^2 v_{22}}, \tag{4.13.5}$$

and

$$\mathrm{Var}(\bar{a}) = \left(\frac{1}{s_1^2 v_{11}} + \frac{1}{s_2^2 v_{22}}\right)^{-1}. \tag{4.13.6}$$

The deviation of $\bar{a}$ from α, divided by its standard error,

$$(\bar{a}-\alpha)\left(\frac{1}{s_1^2 v_{11}}+\frac{1}{s_2^2 v_{22}}\right)^{1/2}, \tag{4.13.7}$$

follows the Behrens distribution with degrees of freedom f_1, f_2, and an angle θ defined by

$$\tan^2\theta = \frac{s_2^2 v_{22}}{s_1^2 v_{11}} \tag{4.13.8}$$

(Yates, 1939; Finney, 1951b). Hence the same distribution may be used in testing the significance of the deviation of a weighted mean, $\bar{a}$, from a theoretical value; alternatively, with d taken as the tabular value for the chosen probability, it allows fiducial limits to α to be placed at

$$d\left(\frac{1}{s^2 v_{11}}+\frac{1}{s^2 v_{22}}\right)^{-1/2}$$

on either side of $\bar{a}$. This neglects information on the distribution of $\bar{a}$ given by the magnitude of $(a-b)$; Fisher (1961a, b) has shown how to take account of the information, which can be important, but no tables exist.

This theory leads to two analogues of Fieller's theorem. For now suppose again that

$$\mu = \frac{\alpha}{\beta}$$

is to be estimated. The ratio of $(a-\mu b)$ to its standard error estimated from

$$\mathrm{Var}(a-\mu b) = s_1^2 v_{11} + \mu^2 s_2^2 v_{22} \tag{4.13.9}$$

follows the Behrens distribution, with f_1, f_2 degrees of freedom and an angle θ given by

$$\tan^2\theta = s_1^2 v_{11}/\mu^2 s_2^2 v_{22}. \tag{4.13.10}$$

As in §4.12, the fiducial limits are the roots of the quadratic equation

$$(a-\mu b)^2 = d^2(s_1^2 v_{11} + \mu^2 s_2^2 v_{22}). \tag{4.13.11}$$

Since d is dependent upon θ, which is in turn a function of μ, no explicit solution of equation (4.13.11) can be given. The solution may be written as

$$m_L, m_U = \left[m \pm \frac{d}{b}\left\{s_1^2 v_{11}(1-g) + m^2 s_2^2 v_{22}\right\}^{1/2}\right] \Big/ (1-g), \tag{4.13.12}$$

where

$$g = \frac{d^2 s_2^2 v_{22}}{b^2}; \tag{4.13.13}$$

numerical evaluation of m_L, m_U, requires interpolation or iteration, since each must have its value of d corresponding to the θ given by equation (4.13.10) when the fiducial limit itself is substituted for μ.

The second analogue of Fieller's theorem requires no iterative calculations, but its statement is more complicated. Suppose that a_1, b_1 and a_2, b_2 are independent pairs of unbiased estimates of α, β as were a, b in §4.12. The variances and covariance for the first pair, $s_1^2 v_{11}$, $s_1^2 v_{22}$, and $s_1^2 v_{12}$, are based upon a single mean square, s_1^2 with f_1 degrees of freedom. The corresponding quantities for a_2, b_2 are based upon an independent mean square s_2^2 with f_2 degrees of freedom, but are in the same ratios as those for a_1, b_1; they may therefore be written $s_2^2 k v_{11}$, $s_2^2 k v_{22}$, and $s_2^2 k v_{12}$, where k is known. The second adaptation of the Behrens distribution can give fiducial limits for a ratio of weighted means of a_1, a_2 and b_1, b_2.

The theorem in §4.12 applies directly to the determination of fiducial limits for a_1/b_1 or a_2/b_2 as estimates of μ. If mean values $\bar{a}, \bar{b}$ are determined by weighting inversely as the variances, so that

$$\bar{a}\left(\frac{1}{s_1^2 v_{11}}+\frac{1}{s_2^2 k v_{11}}\right) = \frac{a_1}{s_1^2 v_{11}} + \frac{a_2}{s_2^2 k v_{11}} \tag{4.13.14}$$

and

$$\bar{b}\left(\frac{1}{s_1^2 v_{22}}+\frac{1}{s_2^2 k v_{22}}\right) = \frac{b_1}{s_1^2 v_{22}} + \frac{b_2}{s_2^2 k v_{22}} \tag{4.13.15}$$

the ratio

$$\bar{m} = \bar{a}/\bar{b} \tag{4.13.16}$$

seems likely to be a more precise estimate of μ than either m_1 or m_2. Now

$$\mathrm{Var}(\bar{a}) = v_{11}\left(\frac{1}{s_1^2}+\frac{1}{k s_2^2}\right)^{-1}, \tag{4.13.17}$$

$$\mathrm{Var}(\bar{b}) = v_{22}\left(\frac{1}{s_1^2}+\frac{1}{k s_2^2}\right)^{-1}, \tag{4.13.18}$$

and the covariance of $\bar{a}, \bar{b}$ is

$$\mathrm{Cov}(\bar{a}, \bar{b}) = v_{12}\left(\frac{1}{s_1^2}+\frac{1}{k s_2^2}\right)^{-1}. \tag{4.13.19}$$

The weights used in forming $\bar{a}, \bar{b}$ are proportional; the variances and covariance of the weighted means therefore preserve the same ratios as those of a_1, b_1 and a_2, b_2. Consideration of $(\bar{a}-\mu\bar{b})$ then gives a quadratic equation for the fiducial limits of $\bar{m}$:

$$(\bar{a}-\mu\bar{b})^2 = d^2(v_{11}-2\mu v_{12}+\mu^2 v_{22}) \Big/ \left(\frac{1}{s_1^2}+\frac{1}{k s_2^2}\right), \tag{4.13.20}$$

where d is a tabular value for f_1, f_2 degrees of freedom and

$$\tan^2\theta = k s_2^2/s_1^2. \tag{4.13.21}$$

If s^2 be defined by

$$s^2 = \left(\frac{1}{s_1^2}+\frac{1}{k s_2^2}\right)^{-1}, \tag{4.13.22}$$

equation (4.13.20) becomes identical with the equality in (4.12.7), except that d replaces t. Hence the solution of equation (4.13.20) may be written in the same form as equation (4.12.3):

$$\bar{m}_L, \bar{m}_U = \left[\bar{m} - \frac{gv_{12}}{v_{22}} \pm \frac{ds}{b}\left\{v_{11} - 2\bar{m}v_{12} + \bar{m}^2 v_{22} - g\left(v_{11} - \frac{v_{12}^2}{v_{22}}\right)\right\}^{1/2}\right] \Big/ (1-g), \tag{4.13.23}$$

where

$$g = \frac{d^2 s^2 v_{22}}{b^2}. \tag{4.13.24}$$

The condition that the variances and covariance of a_2, b_2 should be in the same ratio as those of a_1, b_1 might seem so restrictive as to make the theorem useless. It will be fulfilled, however, by two experiments of the same basic design but possibly different replication – for example, two randomized block experiments for the same treatments, but with different numbers of blocks. The result is therefore useful in the combination of evidence from two assays. If ks_2^2 is large relative to s_1^2, the information on μ provided by a_2, b_2 may be negligible, $\bar{m}$ becomes the same as m_1, d becomes t for f_1 degrees of freedom, and in the limit Fieller's theorem applies.

Generalization of the Behrens distribution to include more than two component variances is conceptually possible, but neither the theory nor tables for a generalized d have been developed. This could form the basis of a further extension to equations (4.13.5)–(4.13.8), and hence to an analogue of Fieller's theorem appropriate to a combination of several assays of similar design. The principle is obvious, and the problem arises again in Chapter 14. A different approach to the combination of estimates from two or more assays is presented in §14.3.

4.14 Fiducial limits in the vitamin D_3 assay

Equation (4.11.5) can be written

$$M - \bar{x}_S + \bar{x}_T = \frac{\bar{y}_T - \bar{y}_S}{b}. \tag{4.14.1}$$

Since $(\bar{x}_S - \bar{x}_T)$ is a constant imposed by the choice of doses, fiducial limits to M may be found by applying Fieller's theorem to the ratio $(\bar{y}_T - \bar{y}_S)/b$ and adding $(\bar{x}_S - \bar{x}_T)$ to the results. From equation (4.12.3), the limits to $(M - \bar{x}_S + \bar{x}_T)$ are

$$\left[M - \bar{x}_S + \bar{x}_T \pm \frac{st}{b}\left\{(1-g)\left(\frac{1}{N_S} + \frac{1}{N_T}\right) + \frac{(M - \bar{x}_S + \bar{x}_T)^2}{\Sigma S_{xx}}\right\}^{1/2}\right] \Big/ (1-g), \tag{4.14.2}$$

where

$$g = \frac{t^2 s^2}{b^2 \Sigma S_{xx}}. \tag{4.14.3}$$

With s^2 taken from equation (4.4.1),

$$g = \frac{(2{\cdot}010)^2 \times 477{\cdot}67}{(18{\cdot}89)^2 \times 207{\cdot}7090}$$
$$= 0{\cdot}0260.$$

An alternative form of calculation is sometimes more convenient; g is the ratio of the tabulated significance point for the variance ratio for 'regression' to the value calculated from the analysis of variance. The variance ratio here has (1, 48) degrees of freedom, whence (Appendix Table II)

$$g = \frac{4{\cdot}04 \times 477{\cdot}67}{74\,088}$$
$$= 0{\cdot}0260.$$

From the form of equation (4.11.6), the limits to $(M + 0{\cdot}1455)$ are

$$\left[0{\cdot}3743 \pm \frac{2{\cdot}010}{18{\cdot}89}\sqrt{\left\{0{\cdot}9740\left(\frac{1}{27} + \frac{1}{28}\right) + \frac{(0{\cdot}3743)^2}{207{\cdot}709}\right\} \times 477{\cdot}67}\right] \Big/ 0{\cdot}9740$$
$$= [0{\cdot}3743 \pm 0{\cdot}6220]/0{\cdot}9740$$
$$= -0{\cdot}2543, 1{\cdot}0229.$$

Therefore

$$M_L = -0{\cdot}3998,$$
$$M_U = 0{\cdot}8774.$$

Equation (4.3.2) gives

$$\log_{10} R_L = \bar{1}{\cdot}0947,$$
$$\log_{10} R_U = \bar{1}{\cdot}2363,$$

whence

$$R_L = 0{\cdot}1244,$$
$$R_U = 0{\cdot}1723.$$

The expression

$$\mathrm{Var}(M) = \frac{s^2}{b^2}\left\{\frac{1}{N_S} + \frac{1}{N_T} + \frac{(M - \bar{x}_S + \bar{x}_T)^2}{\Sigma S_{xx}}\right\}, \tag{4.14.4}$$

frequently quoted as the variance of M, is equation (4.12.9) in the present notation. For these data, approximate fiducial limits to M are obtained by subtracting and adding 2·010 times the standard error (here 0·3135) as in equation (4.12.8). Table 4.14.1 summarizes the results. Evidently g is so small that it could be ignored without harm. Nevertheless, the safer practice of always using Fieller's theorem is little more trouble. Both pairs of limits are presented here for the sake of comparison. Where the concern is with conclusions from assays and not with statistical methodology, quotation of both 'exact' and 'approximate' fiducial limits is to be deprecated as a waste of space on irrelevant and misleading quantities. When g is large, Fieller's theorem should always be used, and values based on any formula for $\mathrm{Var}(M)$ are wrong.

Thus the vitamin content of the cod-liver oil is estimated as 0·1460 units per mg, and the assertion is made that the true potency lies between 85 percent

TABLE 4.14.1 Estimated potency of cod-liver oil (units vitamin D_3 per mg)

	Graphical	Calculated	
		Ignoring g	Fieller's theorem
Potency	0·149	0·1460	0·1460
Lower limit	–	0·1243	0·1244
Upper limit	–	0·1715	0·1723

and 118 percent of this. The fiducial limits are calculated from the internal evidence of a single assay, yet their subsequent use is likely to assume that they measure the agreement to be expected between results of repeated assays of the same test preparation (cf. Finney, 1971, §9.6). Provided that the condition of similarity is fulfilled and that assumptions implicit in the statistical analysis (linearity, homoscedasticity, normality, etc.) are substantially correct, this is justifiable. Potency ought then to be independent of assay technique (§3.4), and the assessment of sampling variation expressed by the fiducial limits ought to have universal validity. The assayist must guard against a too-ready belief that all conditions are satisfied, and that repeated assays will agree within the limits indicated by intra-assay variances. Published experimental verifications are few. Young and Romans (1948) reported satisfactorily consistent potency estimates for 21 insulin samples when each was assayed several times within a few days. Jones (1945) found X-ray and line test assays of vitamin D to agree well during a period of more than three years. Sheps and Munson (1957) proposed a method for taking account of inter-assay variance of M as well as of intra-assay; in a series of androgen assays, they found an important inter-assay component, but did not reconcile this with the general theory of bioassay. The *European Pharmacopoeia* (1971) explicitly counselled that, where possible, precision should be assessed in terms of a simple error mean square calculated from potency estimates from independent assays.

4.15 Data for a symmetric assay

The analysis of the vitamin D_3 assay was complicated because of the unsymmetric design. The coming discussion of the principles of design (especially Chapter 6) may be anticipated by the statement that symmetry is one desirable feature. The simplest symmetric parallel line assays have only two doses of each preparation: the high and low doses of the two preparations have the same difference on the logarithmic scale, and the total number of subjects is divided equally between doses.

Table 4.15.1 contains data from an assay of oestrone using 7 litters of 4 ovariectomized female rats each. Each rat was injected daily with one of the four experimental doses, 0·2 μg and 0·4 μg of the standard oestrone and 0·0075 ml and 0·015 ml of the test preparation. The response was the weight of the uterus, expressed as mg per 100 g body weight and measured at a fixed number of days after treatment of the animals (Bülbring and Burn, 1935). This method of

TABLE 4.15.1 Weights of uteri of ovariectomized rats, in mg per 100 g body weight

Litter	Daily dose				Totals
	Oestrone		Test preparation		
	0·2 μg	0·4 μg	0·0075 ml	0·015 ml	
I	54	152	61	92	359
II	49	71	74	63	257
III	51	112	51	(87)	301
IV	(50)	58	60	102	270
V	81	102	(82)	120	385
VI	63	(111)	83	105	362
VII	126	(133)	83	108	450
Totals	474	739	494	677	2384
Means	67·7	105·6	70·6	96·7	

Five responses are shown in parentheses, for reasons explained in §4.21.

adjusting for the sizes of the subjects will be criticized in §12.6; here records of body weights are no longer available and therefore covariance analysis could not be tried. If litters were to differ in mean uterine weight, as might be expected, the precision of the assay could have been adversely affected by inter-litter variation. This was avoided by adopting a common device, a randomized block design; one animal from each litter, selected at random, was assigned to each of the four doses. For simplicity, some liberties have been taken with the data, as explained in §4.21.

The reader should draw a dose–response diagram for the mean responses in Table 4.15.1. A convenient dose scale, using logarithms to base $\sqrt{2}$, makes $(x-\bar{x})$ equal to -1 for the lower, $+1$ for the upper dose of either preparation. The two preparations may have different origins on the x scale, and a simple choice is that which makes the two lower doses have the same scale point. Parallel regression lines drawn by eye in this diagram give a rough estimate of potency. One version of the diagram had lines 0·18 apart in a direction parallel to the axis of x. Hence

$$M = \bar{x}_S - \bar{x}_T - 0{\cdot}18,$$

and, using a formula similar to equation (4.3.2),

$$\log_{10} R = \log_{10} 0{\cdot}2 - \log_{10} 0{\cdot}0075 - 0{\cdot}18 \times \tfrac{1}{2} \times \log_{10} 2, \qquad (4.15.1)$$

since symmetry makes $(\bar{x}_S - \bar{x}_T)$ equal to the difference in x values for corresponding doses. Therefore

$$R = \frac{0{\cdot}2}{0{\cdot}0075} \text{ antilog } \bar{1}{\cdot}973 \qquad (4.15.2)$$

$$= 25,$$

and 1 ml of the test preparation is estimated to contain 25 μg oestrone.

4.16 Analysis of variance

Table 4.16.1 shows the analysis of variance of the 28 entries in the body of Table 4.15.1. The subdivision of the total sum of squares into components 'between doses', 'between litters', and 'error' follows the usual procedure for a randomized block design: for example, the sum of squares between doses is

$$(474^2 + 739^2 + 494^2 + 677^2 - 7 \times 202\,981)/7.$$

Subdivision of the dose component into squares for three separate degrees of freedom can be effected by the same steps as were used to give lines 4, 5, 6 in Table 4.4.1. For a symmetric design, orthogonal contrasts as illustrated in Table 4.16.2 are more convenient. The dose totals are multiplied in turn by each row of coefficients and summed to give the sums of products in the last column, each of which will be denoted by the letter L with a distinguishing subscript. These sums are squared and divided by the divisors shown; the quotients are the components for Table 4.16.1.

TABLE 4.16.1 Analysis of variance for Table 4.15.1

Adjustment for mean		202 981	
Nature of variation	d.f.	Sum of squares	Mean square
Preparations	1	63	63
Regression	1	7 168	7168
Parallelism	1	240	240
Between doses	3	7 471	
Between litters	6	7 069	
Error	18	7 165	398·1
Total	27	21 705	

TABLE 4.16.2 Coefficients of orthogonal contrasts for the (2, 2) design, applied to Table 4.15.1

Dose	S_1	S_2	T_1	T_2	Divisor	Sum
Response total	474	739	494	677	28	2384
Preparations (L_p)	−1	−1	1	1	28	−42
Regression (L_1)	−1	1	−1	1	28	448
Parallelism (L_1')	1	−1	−1	1	28	−82

Coefficients such as those in Table 4.16.2 lead to a subdivision of the sum of squares for doses *only* if:

(i) Each row of coefficients represents a *contrast* amongst the individual responses; that is to say, if each sum in Table 4.16.2 is written at length in terms of individual responses, the set of coefficients for these responses adds to zero;

(ii) Every pair of contrasts is *orthogonal*; that is to say, if each contrast is expressed in terms of individual responses, the products of corresponding coefficients for any pair add to zero.

Because each dose total in a symmetric design contains the same number of individual responses, these conditions also apply to the coefficients of dose totals. Since

$$-1 + 1 - 1 + 1 = 0$$

the line for 'Regression' represents a contrast, and since

$$(-1) \times 1 + 1 \times (-1) + (-1) \times (-1) + 1 \times 1 = 0$$

the contrasts for 'Regression' and 'Parallelism' are orthogonal. The divisors are calculated as the sums of the squares of the coefficients of individual responses; for 'Regression',

$$7 \times (-1)^2 + 7 \times 1^2 + 7 \times (-1)^2 + 7 \times 1^2 = 28.$$

If the variance per response is σ^2, the variance of a contrast value is σ^2 multiplied by this 'divisor'. A consequence of these conditions is that the squares for the separate contrasts must add to the sum of squares for doses; here,

$$\frac{(-42)^2}{28} + \frac{448^2}{28} + \frac{(-82)^2}{28} = 7471.$$

Any sum of squares of deviations can be subdivided into single squares for mutually orthogonal contrasts, in number equal to the degrees of freedom, in an unlimited number of ways. The set in Table 4.16.2 is chosen as peculiarly relevant to the object of the analysis of variance. The three squares are exactly the same as would have been obtained from the general method of §4.4. The first contrast gives the difference between totals for the two preparations; the second has $(x - \bar{x})$ as its coefficients, so that the sum is ΣS_{xy}; and the third gives the difference between values of S_{xy} for the preparations. Bliss and Marks (1939a, b) showed the advantages of such coefficients in the analysis of assays (cf. Chapter 5).

4.17 Validity tests

As in §§4.7 and 4.8, the mean squares for preparations and parallelism must be compared with the error mean square. Neither is significantly large, so that on these counts the validity of the assay need not be doubted. The great flaw in the design is that it gives no test for linearity, since only two points on each response curve are studied. A genuinely non-linear regression might manifest itself as non-parallelism, at least if the doses had been so unsuccessfully chosen as to give also a large difference between preparations (Fig. 4.10.3). On the other hand, if doses of T were so chosen that they were almost equal in effect to corresponding doses of S, the fitted lines would appear satisfactorily parallel, even though the true relation was far from linear (Fig. 4.10.1). As pointed out in §4.10, this will not bias the estimation of potency appreciably, but may upset the assessment of precision. Assays should usually be planned to include at least three doses of each preparation (§6.10), unless the material under assay is so well understood as to remove all fear of non-linearity. No assay designed in randomized blocks

allows a test of scedasticity, but this is usually less important once an assay technique is well established.

4.18 Potency estimation and precision

The construction of Table 4.16.2 makes clear that the contrast labelled L_p is the difference in total responses for the two preparations. Hence

$$\begin{aligned} \bar{y}_T - \bar{y}_S &= L_p/14 \\ &= -42/14 \\ &= -3{\cdot}00. \end{aligned} \tag{4.18.1}$$

Moreover, the divisor and sum for the regression contrast, L_1, are ΣS_{xx} and ΣS_{xy} respectively, and therefore, by equation (4.11.1),

$$\begin{aligned} b &= L_1/28 \\ &= 448/28 \\ &= 16{\cdot}00. \end{aligned} \tag{4.18.2}$$

Equation (4.11.5) now gives

$$M = \bar{x}_S - \bar{x}_T - 0{\cdot}1875,$$

and, by equation (4.15.1),

$$\begin{aligned} R &= \frac{0{\cdot}2}{0{\cdot}0075} \text{ antilog } (-0{\cdot}1875 \times 0{\cdot}1505) \\ &= 25{\cdot}0. \end{aligned}$$

Fiducial limits may be found by Fieller's theorem, using the error mean square in Table 4.16.1:

$$s^2 = 398{\cdot}1$$

as the variance per response. By the well-known elementary property of a linear function of independent observations, the variance of any contrast can be written as the variance per response multiplied by the sum of the squares of the coefficients of individual responses in the contrast. The rule by which the divisors in Table 4.16.2 are calculated shows this to be expressible as

$$\text{Var}(L) = s^2 \times \text{Divisor} \tag{4.18.3}$$

for any contrast, or

$$\text{Var}\left(\frac{L}{\text{Divisor}}\right) = \frac{s^2}{\text{Divisor}}. \tag{4.18.4}$$

Consequently, from equations (4.18.1) and (4.18.2),

$$\text{Var}(\bar{y}_T - \bar{y}_S) = s^2/7 \tag{4.18.5}$$

and

$$\text{Var}(b) = s^2/28. \tag{4.18.6}$$

By equation (4.14.3)

$$\begin{aligned} g &= \frac{(2{\cdot}101)^2 \times 398{\cdot}1}{(16{\cdot}00)^2 \times 28} \\ &= 0{\cdot}2452, \end{aligned}$$

or, from Table 4.16.1 and the alternative method in §4.14,

$$\begin{aligned} g &= \frac{4{\cdot}41 \times 398{\cdot}1}{7168} \\ &= 0{\cdot}2449, \end{aligned}$$

a value that is arithmetically slightly less accurate. From equation (4.14.2), the fiducial limits to $(M - \bar{x}_S + \bar{x}_T)$ are

$$\begin{aligned} &\left[-0{\cdot}1875 \pm \frac{2{\cdot}101}{16{\cdot}00}\left\{\left(\frac{0{\cdot}7548}{7} + \frac{0{\cdot}1875^2}{28}\right) \times 398{\cdot}1\right\}^{1/2}\right] \Big/ 0{\cdot}7548 \\ &= [-0{\cdot}1875 \pm 0{\cdot}8653]/0{\cdot}7548 \\ &= -1{\cdot}3948, 0{\cdot}8980. \end{aligned}$$

Therefore

$$\begin{aligned} R_L &= \frac{0{\cdot}2}{0{\cdot}0075} \text{ antilog } \bar{1}{\cdot}7901 = 16{\cdot}5, \\ R_U &= \frac{0{\cdot}2}{0{\cdot}0075} \text{ antilog } 0{\cdot}1351 = 36{\cdot}4. \end{aligned}$$

Thus the potency is estimated to be 25·0 μg per ml, with fiducial limits at 16·5 μg and 36·4 μg per ml.

4.19 Constraints of design

The design of any experiment determines the character of the proper statistical analysis. In the oestrone assay, litter-mate control was adopted so that differences between litters would not affect the estimate of potency or the assessment of its precision; to analyze the data of Table 4.15.1 ignoring the litter classification would be logically wrong and possibly very misleading.

An assayist might find it convenient to put all identically treated animals into one cage (or, in a microbiological assay, to put all tubes of the same dose in adjacent positions in the incubator). Rarely is this advisable, for it *confounds* (§9.2) differences between doses with differences between cages: the several animals in one cage are not true replicates of the treatment for comparison with differently treated animals in another cage. Interaction between the animals in one cage, such as competition for food, may make individual responses different from what they would have been had all animals been caged separately, so producing a variance between cages different from that within cages. Anyone who

analyzes the responses without regard to cage differences is in effect asserting that these differences are negligible, and that he may legitimately assess the precision of potency estimate from variation *within* cages in spite of the fact that the dose contrasts used are made *between* cages; the experiment itself can provide no test of the validity of this assumption, unless each dose group is spread over two or more cages. Even though the animals were caged individually, the same difficulty would arise if all cages for one dose were placed close together in the animal house. From extensive experience, Emmens (1948, §13.5) wrote: 'There has been in biological work a considerable tendency to ignore the possibility of differences in reaction due to animals being caged in distinct groups and it seems to have been tacitly assumed that variation between cages must be negligible. It must be a rarely designed animal house in which conditions are so uniform that this assumption can be justified, and in the light of our knowledge that a variety of responses are influenced by health, temperature, light, feeding and many other factors, it would always seem worth while so to arrange our preliminary trials that the contributions of these factors to differences in the location of test objects may be examined'.

The statistical analysis of an experiment is a small part of the total labour, and its costs should not influence the choice of design (§6.6). The arrangement of subjects in cages, the randomization of order in an animal house, or the randomization of order of testing, may bring the convenience or even the practicability of an experiment into conflict with the ideal statistical conditions. Complex designs (Chapters 9 and 10) sometimes enable the statistician to overcome these difficulties, but the very complexity of a design can also make it impossible of application. Individual caging of subjects, or a complicated arrangement of tests, may so much increase the risk of gross mistakes or the costliness of an experiment as to make it completely impracticable. The statistician must recognize that these situations do arise, and must be prepared for some compromise with the exigencies of experimentation. He will need to make clear to the assayist the price that must be paid for the use of a statistically inferior design: loss of precision or, more serious, conclusions whose validity rests upon an untestable assertion about the unimportance of certain sources of variation. If the assayist is satisfied that these disadvantages do not outweigh the advantages of the design, the statistician's responsibility is ended.

Ideally, animals in an assay are caged individually, caged in groups corresponding to one of the uninformative classifications of the experiment (such as litters), or caged in some new groupings, orthogonal with all others, for which a sum of squares can be isolated in the analysis of variance. If individual caging is impracticable, and the assayist is reluctant to cage together animals that are being differently treated (possibly for the good reason that they would affect one another), he should at least aim at dividing each dose group between two or more cages. The analysis of variance will then show separate residual mean squares between and within cages; if the first is significantly larger than the second, it must be used as the basic s^2 for subsequent calculations.

When faced with the results of an assay for the detailed design of which he

was not responsible, the statistician must discover exactly how the experiment was arranged and conducted. Bitter experience will teach him how easily an experimenter may fail to mention the existence of a constraint, because of failure to realize its relevance to the statistical analysis: in some instances, the appearance of the data may arouse the suspicions of an alert statistician, but in others only the most careful discussion of the experiment with the person responsible for its execution will elicit information that vitally affects the statistical analysis.

4.20 Heterogeneous deviations from linearity

Deviations from linearity of regression were discussed in §4.6. If an assay in which several doses of each preparation were included shows significant non-linearity, inspection of the dose–response diagram may show either of two situations. Systematic deviation of the points from the calculated straight lines may indicate that the true regressions are curved (cf. Fig. 4.10.1); the linear regression model is wrong, and must be rejected in favour of a different metametric transformation or a different method of analysis. Alternatively, the points may show considerable scatter about the lines yet appear completely erratic in their deviations; this may be a manifestation of an unusually complicated dose–response relationship, but often a more plausible explanation is heterogeneity of the batches of subjects at different doses. Even though the true regression be linear, if subjects were not assigned at random to doses (or if indeed the batches were knowingly made up from different sources), the deviations of mean responses from the regression lines will be greater than is predicted from variations within batches. Similar trouble may arise if different doses have to be tested on different occasions, and experimental conditions change between occasions.

If the assayist is prepared to accept heterogeneity between the dose groups, rather than a very complex regression curve, as the explanation of erratic deviations from linearity, the mean square for deviations from linearity may be used as the estimated variance per response in all subsequent tests and assessments of fiducial limits. Randomized allocation of groups, occasions, or other classifications, to doses is essential. For example, if only one dose per day can be tested, and the doses are used in systematic order on successive days, any secular trend in experimental conditions will bias estimation of the regression coefficient; a random order of doses will ensure that deviations about regression lines give a valid estimate of the random errors of experimentation. In an assay of an insecticide, successive batches of insects from a single culture might show a steady trend in sex-ratio; random allocation of batches to doses ensures that the estimation of potency and the assessment of precision are unbiased by any correlation of sex-ratio with response (Bliss, 1939; Murray, 1937). When the only alternative is to reject data as worthless, the temptation to accept the mean square for deviations from linearity as s^2 is strong. The critical reader will appreciate the need for restraint. To assume randomness when no random element has been incorporated into the design is a great risk. For validity of such a variance estimate, theory requires the number of subjects to be the same at each dose, but in practice slight inequalities do not matter.

When only two or three doses of each preparation are tested, discrimination between the two types of significant deviation from linearity is impossible. This might appear opposed to the recommendations of §6.10 on the number of dose levels. In reality, the right course is almost always to choose an assay design that makes proper randomness consistent with inevitable restrictions on experimental technique, thus avoiding any need for an estimate of variance based upon deviations from linearity. If this is not practicable, the case for four or more doses of each preparation is strengthened. The data of Table 4.2.1 give only 3 d.f. for the linearity mean square; if fiducial limits had to be based on this, they would have suffered from the imprecision in the estimate of variance. An illustration of the use of the linearity mean square as s^2 occurs in §16.2. In assays using quantal responses, the same problem may arise: the heterogeneity factor (§18.1) has the same function, and the same unsatisfactory basis, as the variance estimate just described.

It is important to avoid any automatic rule of rejecting assays on account of non-linearity or other aspects of statistical invalidity. As Humphrey *et al.* (1953) have emphasized, a rule based solely on individual significance tests would merely result in the most precise assays being rejected! A truly linear regression is a rarity, and to penalize all assays in which high precision detects non-linearity is folly. To formulate an ideal policy is difficult, as significant non-linearity at least indicates that precision is less good than the error mean square suggests. Humphrey's practice seems somewhat less desirable than the use of a different mean square for s^2, but the question deserves closer study by those concerned with large numbers of related routine assays.

4.21 Missing values

Even in the most carefully conducted experiment, an accident or unforeseeable circumstance may cause the loss of a subject and so destroy the symmetry of a design. Restrictions on the subjects available for use may even prevent adoption of a symmetric design. In the assay to which Table 4.15.1 refers, the responses recorded for five of the litters relate only to three rats; whether this was because only three female litter-mates were available or because animals were lost during the experiment is not now known. The missing records correspond to the positions marked by parentheses in Table 4.15.1. If the reason was that five litters had only three females, the design adopted was about the best that could be contrived; certainly it was preferable to omitting one dose entirely or from each of the five litters.

When observations are missing from a randomized block or more complicated design, special procedures are required to prevent the gaps inducing biased comparisons of dose means. This is because the orthogonality is destroyed: if the coefficients of Table 4.16.2 are applied to a new version of Table 4.15.1 that contains only the 23 genuine responses, neither the contrast nor the orthogonality condition is satisfied. Inspection of the data indicates large differences between litters, so that simple averaging of columns would give an unfair representation. Litter IV, for example, gave low results, and therefore the average of the

responses of the other six litters for the lower dose of S would give too high a value relative to other doses.

One way of overcoming the difficulty is to calculate from the genuine records values that, when inserted in the empty spaces of the table, will remove any distortion in the means. The general procedure is to use symbols $y_1, y_2, y_3, \ldots$ to represent the missing values, to perform an analysis of variance of all the data in terms of these symbols, to express the error sum of squares as a quadratic function of the unknowns, and then to determine $y_1, y_2, y_3, \ldots$ by the condition that the error sum of squares shall be a minimum. This is a standard adaptation of the analysis of variance. When a single value must be calculated in a randomized block design, it leads to the formula

$$y = \frac{rR + cC - G}{(r-1)(c-1)}; \qquad (4.21.1)$$

r is the number of rows (here litters) in the table of results and R the total of all genuine data in the same row as the missing entry,(*) c is the number of columns (here doses) and C the total of all genuine data in the same column as the missing entry, and G is the total of all the data. Equation (4.21.1) represents a compromise between the average of all other entries in the same row and the average of all other entries in the same column. When more than one entry is missing, the formula may be applied iteratively. Values are guessed for all except the first, and the formula is used to calculate the first; the result, together with all guessed values except that for the second missing entry, is then used in a calculation of the second, and so on until all have been calculated. The process is repeated so as to revise the first, second, . . . values with the aid of the results of the first set of calculations, and the iteration is continued until two successive cycles agree closely. The final values are independent of the initial guesses.

In the oestrone assay, y_1, y_2, y_3, y_4, y_5 represent the missing responses in Litters III, IV, V, VI, VII, and the formula for iteration is

$$y = (7R + 4C - G)/18.$$

The process may start with any values, for example 100 for each of y_2, y_3, y_4, y_5. Remembering that R, C, G must include every response except the one currently being recalculated, the formula gives

$$y_1 = (7 \times 214 + 4 \times 590 - 2321)/18 = 85$$

and then

$$y_2 = (7 \times 220 + 4 \times 424 - 2306)/18 = 52$$

and so on. The complete iteration is

(*)A momentary use of R in a sense different from relative potency should not confuse the reader; it will occur only when missing data have to be discussed.

	y_1	y_2	y_3	y_4	y_5
Start	–	100	100	100	100
Cycle 1	85	52	84	105	132
Cycle 2	87	50	82	111	133
Cycle 3	87	50	82	111	133

Any alternative starting values would lead to the same results. Although iteration could be continued to establish one or more decimal digits, in practice it is pointless to evaluate to greater accuracy than that of the recorded responses.

The possibility of calculating 'missing values' in this manner is no justification for careless experimentation, either in choice of design or in failure to make complete records. Though the calculated values are estimates of the responses that would have been found in a complete experiment, they do not create information by some statistical trick. Their primary function is to eliminate bias in the comparison of means. Table 4.15.1 was in fact constructed by insertion of the values just calculated, and in §§4.16–4.18 the assay was analyzed without comment. However, in an analysis of variance using calculated entries as though they were genuine, the error mean square is an unbiased estimate of the variance per response only if the number of degrees of freedom for error is reduced by the number of missing entries inserted. The loss of information is felt when mean responses for different doses are compared. In particular, though the functions of mean responses used in the formation of $(\bar{y}_T - \bar{y}_S)$ and b are unbiased, their variances are increased.

4.22 Approximate analyses for missing values

The simplest adjustment that can be made to take account of the missing values in the oestrone assay is merely to use the correct degrees of freedom for error, $(18 - 5)$. From Table 4.16.1, the variance should be estimated as

$$s^2 = 7165/13$$
$$= 551{\cdot}2.$$

Repetition of the calculations in §4.18 with this value for s^2 and 13 d.f. gives fiducial limits 14·3 μg and 40·6 μg per ml.

Had only one response been missing, this would have been fairly satisfactory. With this high proportion of 5 missing out of 28, the approximation is not very good. A further modification will commonly give conclusions near enough to those from the exact analysis in §4.23. This consists in obtaining L_p, L_1 from the estimated values in §4.21, using equations (4.18.1) and (4.18.2), but writing

$$\text{Var}(L_p/14) = s^2\left(\frac{1}{6}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}\right)\Big/4 \qquad (4.22.1)$$
$$= 0{\cdot}175\, s^2$$

$$\text{Var}(L_1/14) = s^2\left(\frac{1}{6}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}\right)\Big/16 \qquad (4.22.2)$$
$$= 0{\cdot}043\,75 s^2;$$

these are based upon assigning to the mean for each dose an effective variance $s^2/6$ or $s^2/5$, instead of $s^2/7$, because only 6 or 5 genuine response measurements are available. Recalculation from this point on exactly as in §4.18 gives

$$g = 0{\cdot}4395,$$

and then a potency estimate of 25·0 μg per ml but with fiducial limits, obtained by using equations (4.22.1) and (4.22.2) in Fieller's theorem (assuming zero covariance), at 12·8 μg and 44·1 μg per ml.

The limits found by this approximation are a little narrower than the correct ones in §4.23, but are close enough for practical purposes. Had only one or two entries been missing, the approximation would have been still more satisfactory. On the other hand, had the missing entries left the design even more unbalanced (e.g. three missing from the S_1 dose and two from T_2), it might not have been good enough. Yates's method (1933) for calculating the variances of contrasts between means after adjustment for missing entries could be adapted for use here. In any assay for which neither method in the present section is good enough, however, the full analysis illustrated in §4.23 is today so little more laborious that it should be used.

4.23 Exact analysis for missing values

The contrasts expressed by the coefficients in Table 4.16.2 are mutually orthogonal only when true responses have been recorded for all subjects, as was originally assumed for Table 4.15.1. The calculations in §4.21 provide for the five empty spaces values that are functions of the other 23. If these were written in full, it would become apparent that L_p, L_1, L_1' are still contrasts but are no longer orthogonal and no longer have variances given by equation (4.18.3). Exact analysis could proceed on these lines, with examination of the contrasts for L_p, L_1; determination of their variances and covariance as linear functions of the 23 responses, and application of Fieller's theorem to L_p/L_1.

This would be tedious, and would require new examination of contrasts for each new configuration of missing values encountered. General methods of fitting constants or solving least squares problems for linear models can be applied; unless care is exercised in formulation of the problem and the parameters, the arithmetic can become heavy, involving inversion of a fairly large matrix. An easier procedure to adopt as standard for bioassay is that based upon multiple linear regression. The great advantage is that the computing routine is widely known, standard programs are available on almost all scientific computers, and even on desk calculators an operator with a little experience of statistical calculation will need minimal special instruction. Although described here in terms of a very simple assay, the method readily adapts to other situations.

The key lies in the realization that all the information required for validity tests and for potency estimation is contained in a set of contrasts between doses. In the oestrone assay, these are three, and they can be formally identified with a set of three partial regression coefficients. Introduce three *dummy variates*, x_1, x_2, x_3, defined to take the following values at the four dose levels:

	S_1	S_2	T_1	T_2
x_1 =	−1	−1	1	1
x_2 =	−1	1	−1	1
x_3 =	1	−1	−1	1

The variates have the same values as the coefficients in Table 4.16.2, and x_2 is the $(x - \bar{x})$ of the earlier analysis. For complete data with no missing entries, the regressions on x_1, x_2, x_3 would give the sums in the last column of Table 4.16.2 as the various S_{xy}. Because of orthogonality of the dummy variates, the three squares for preparations, regression and linearity in Table 4.16.1 would then be computed as the squares attributable to the separate regressions.

The method may be applied to the genuine data of Table 4.15.1, after omission of bracketed entries, and is numerically equivalent to other methods mentioned above. An analysis of variance and covariance for x_1, x_2, x_3, y is made, all distinctions between columns (doses) being dropped since these are taken account of by the independent variates. Sums of squares and products are calculated for the 22 d.f. between the 23 responses, and components with 6 d.f. for litter differences are subtracted. Some litters contain four responses and some only three. For y, for example, the sum of squares for litters is

$$\tfrac{1}{4}(359^2 + 257^2) + \tfrac{1}{3}(214^2 + 220^2 + 303^2 + 251^2 + 317^2) - \frac{1921^2}{23} = 4786;$$

similarly, the sum of products of x_2, y for litters is

$$\tfrac{1}{3}(-1 \times 214 + 1 \times 220 + 1 \times 303 - 1 \times 251 - 1 \times 317 - \frac{(-1) \times 1921}{23} = 6{\cdot}81.$$

The reader should have no difficulty in checking the details of the analysis in Table 4.23.1.

The within-litter regression coefficients, b_1, b_2, b_3, are the solutions of

$$\left.\begin{aligned} \tfrac{64}{3}b_1 + \tfrac{4}{3}b_2 - \tfrac{4}{3}b_3 &= -\tfrac{22}{3}, \\ \tfrac{4}{3}b_1 + \tfrac{64}{3}b_2 + \tfrac{4}{3}b_3 &= \tfrac{1006}{3}, \\ -\tfrac{4}{3}b_1 + \tfrac{4}{3}b_2 + \tfrac{64}{3}b_3 &= -\tfrac{116}{3}. \end{aligned}\right\} \qquad (4.23.1)$$

In these equations, exact numerical values have been inserted (instead of their decimal expressions in Table 4.23.1) because they are here so simple in form. Solution requires inversion of the 3 x 3 matrix of coefficients:

$$\frac{1}{3}\begin{pmatrix} 64 & 4 & -4 \\ 4 & 64 & 4 \\ -4 & 4 & 64 \end{pmatrix}. \qquad (4.23.2)$$

TABLE 4.23.1 Analysis of variance and covariance for exact analysis of Table 4.15.1

Adjustments for means		0·0435	0·0435	0·0435	0·0435	83·52	−83·52	83·52	160 445
Nature of variation	d.f.	$S_{x_1x_1}$	$S_{x_1x_2}$	$S_{x_1x_3}$	$S_{x_2x_3}$	S_{x_1y}	S_{x_2y}	S_{x_3y}	S_{yy}
Between litters	6	1·6232	0·2898	0·2898	−0·2898	6·81	−2·81	62·15	4 786
Within litters	16	21·3333	1·3333	−1·3333	1·3333	−7·33	335·33	−38·67	12 652
Total	22	22·9565	1·6232	1·0435	−1·0435	−0·52	332·52	23·48	17 438
Adjustments for means		0·0435	0·0435	0·0435	0·0435				
Nature of variation	d.f.	$S_{x_1x_1}$	$S_{x_2x_2}$	$S_{x_3x_3}$					
Between litters	6	1·6232	1·6232	1·6232					
Within litters	16	21·3333	21·3333	21·3333					
Total	22	22·9565	22·9565	22·9565					

By standard procedures, such as are well described by Searle (1966), the inverse matrix is found to be

$$\mathbf{V} = \frac{3}{952}\begin{pmatrix} 15 & -1 & 1 \\ -1 & 15 & -1 \\ 1 & -1 & 15 \end{pmatrix}. \qquad (4.23.3)$$

Those unfamiliar with matrices may note that, for an initial matrix that is symmetric, the sum of products of corresponding rows is unity; for example, from the second rows in (4.23.2) and (4.23.3)

$$(-1 \times 4 + 15 \times 64 - 1 \times 4)/952 = 1.$$

Sums of products of non-corresponding rows are zero; from the second row of (4.23.2) and the third of (4.23.3)

$$(1 \times 4 - 1 \times 64 + 15 \times 4)/952 = 0.$$

To be able to take advantage of simple expressions in fractions is unusual. In many practical situations, the inversion would have to be done decimally; many digits would need to be retained in **V** as an aid to checking and as a guard against arithmetical inaccuracy arising because differences between nearly equal quantities are evaluated. Matrix inversion is an arithmetical operation especially well suited to high-speed computers, and appropriate routines are always included in standard programs for regression and covariance analyses; even these can be inaccurate for large matrices. A desk calculator can be used, but the labour of inverting a 6 x 6 or larger matrix is then liable to be very heavy.

The element in row i, column j of **V** may be denoted v_{ij}. The regression coefficients are then obtained as sums of products of the quantities on the right of equations (4.23.1) with each row of **V** in succession. Thus

$$b_i = (-22v_{i1} + 1006v_{i2} - 116v_{i3})/3, \qquad (4.23.4)$$

using the obvious exact fractional values. Hence

$$b_1 = (-22 \times 15 - 1006 - 116)/952$$
$$= -1{\cdot}5252,$$
$$b_2 = 15{\cdot}9958,$$
$$b_3 = -2{\cdot}9076.$$

The regression accounts for a sum of squares

$$(-22b_1 + 1006b_2 - 116b_3)/3 = 5488. \qquad (4.23.5)$$

Regression on log dose (x_2) alone would account for a component of this obtained by omitting x_1, x_3 and repeating the calculations, or more directly from the familiar formula

$$(335{\cdot}3333)^2/21{\cdot}3333 = 5271.$$

The difference, with 2 d.f., is a composite test of preparations and parallelism, which cannot be separated completely because of non-orthogonality. As Table 4.23.2 shows, the two together are too small to occasion any concern; a test of either could be made by omitting x_2 or x_3 and finding out how much of the 5488 is left when a regression on x_1 and the other is formed, but the portions so obtained would not be independent and additive (cf. Table 2.8.3). Note that the error sum of squares is equal to that in Table 4.16.1; the procedures that gave the two are algebraically identical, though a small discrepancy appears because the arithmetic of estimating missing values in §4.21 was carried only to the nearest integers. Table 4.23.2, however, gives unbiased validity tests.

TABLE 4.23.2 Final analysis of variance for exact analysis of Table 4.15.1

Adjustment for mean		160 445	Mean square
Nature of variation	d.f.	Sum of squares	
Regression on log dose	1	5 271	
Preparations and Parallelism	2	217	108
Doses	3	5 488	
Litters, ignoring doses	6	4 786	
Error	13	7 164	551·1
Total	22	17 438	

The definitions show that the regression coefficient on x_2 estimates the regression of response on log dose, the quantity usually called b. Similarly, the regression coefficient on x_1 is an estimate of one-half the difference in mean responses for the two preparations, the quantity usually called $(\bar{y}_T - \bar{y}_S)$. The method has ensured that these estimates are adjusted for non-orthogonality; they differ from those in §4.21 only because of the rounding of decimal digits in that section. Hence M is taken as

$$M = \bar{x}_S - \bar{x}_T - \frac{2 \times 1{\cdot}5252}{15{\cdot}9958}$$
$$= \bar{x}_S - \bar{x}_T - 0{\cdot}1907. \qquad (4.23.6)$$

In order to construct the fiducial limits for M, the variances and covariance of $2b_1$ and b_2 are needed. They are obtained from the matrix **V** as

$$\left.\begin{aligned} \mathrm{Var}(2b_1) &= 4s^2v_{11} = 45s^2/238, \\ \mathrm{Var}(b_2) &= s^2v_{22} = 45s^2/952, \\ \mathrm{Cov}(2b_1, b_2) &= 2s^2v_{12} = -3s^2/476. \end{aligned}\right\} \qquad (4.23.7)$$

Hence

$$g = \frac{(2{\cdot}160)^2 \times 45 \times 551{\cdot}1}{952 \times (15{\cdot}9958)^2}$$

$$= 0{\cdot}4750. \qquad (4.23.8)$$

Fieller's theorem gives fiducial limits for $(M - \bar{x}_S + \bar{x}_T)$; from equations (4.23.7) they are

$$\left[-0{\cdot}1907 + \frac{0{\cdot}4750 \times 6}{45} \pm \frac{2{\cdot}160}{15{\cdot}9958}\{(180 - 2 \times 0{\cdot}1907 \times 6 + 0{\cdot}1907^2 \times 45 - 0{\cdot}4750 \times 179{\cdot}2) \times 551{\cdot}1/952\}^{1/2}\right] \Big/ 0{\cdot}5250$$

$$= (-0{\cdot}1907 + 0{\cdot}0633 \pm 0{\cdot}9974)/0{\cdot}5250$$

$$= -2{\cdot}1425, 1{\cdot}6571.$$

Once again from equation (4.15.1),

$$R = \frac{0{\cdot}2}{0{\cdot}0075} \text{antilog } \bar{1}{\cdot}9713 = 25{\cdot}0,$$

and similarly

$$R_L = \frac{0{\cdot}2}{0{\cdot}0075} \text{antilog } \bar{1}{\cdot}6776 = 12{\cdot}7,$$

$$R_U = \frac{0{\cdot}2}{0{\cdot}0075} \text{antilog } 0{\cdot}2494 = 47{\cdot}4.$$

The potency estimate could differ from that in §4.22 only because more digits were retained in the present calculations; to the accuracy that may reasonably be reported, it is identical with that obtained earlier. The widening of the limits to 12·7 μg and 47·4 μg per ml represents the effect of failure to adjust the first analysis adequately on account of the missing entries in Table 4.15.1.

5

Symmetric dose - structure for parallel line assays

5.1 The general structure

Chapter 4 has illustrated both the advantages of symmetry in the design of biological assays and the inadequacy of the simplest symmetric assay in respect of validity tests. A more general symmetric dose-structure, termed a (k, k)-point assay, has k doses of each preparation such that successive doses of either preparation bear a ratio D to one another ($D > 1$), with n subjects at each dose. The total number of subjects,

$$N = 2nk, \qquad (5.1.1)$$

includes nk for each preparation. Before discussion of the general analysis, an example of a (3, 3) assay is presented.

5.2 An assay of vitamin B_{12}

Emery *et al.* (1951) reported an assay conducted according to the principles of the dose–response relation studied in §3.9. This assay has $k = 6$, $N = 36$, and $D = 1{\cdot}5$. By taking the dose metameter

$$x = \log_{1{\cdot}5}(z/z_c), \qquad (5.2.1)$$

where z_c is the central dose (1·2 ng of S, 6 units of T), the values of x are made -1, 0, 1 for each preparation. Table 5.2.1 contains the response metameters obtained from equation (3.9.6). Fig. 5.2.1, showing mean responses plotted against x, suggests both linearity and parallelism.

TABLE 5.2.1 Response metameters in an assay of vitamin B_{12}

Standard preparation (ng/tube)			Test preparation (units/tube)		
0·8 S_1	1·2 S_2	1·8 S_3	4 T_1	6 T_2	9 T_3
0·96	1·06	1·17	0·91	1·09	1·15
0·91	1·07	1·14	0·93	1·04	1·15
0·92	0·99	1·14	0·98	0·97	1·14
0·76	0·86	1·13	0·96	1·06	1·16
1·03	1·06	1·13	0·89	1·04	1·10
0·93	1·02	1·15	1·01	1·02	1·15
5·51	6·06	6·86	5·68	6·22	6·85

The analysis of variance, Table 5.2.2, subdivides the total sum of squares into components between doses and within doses (error); the assay has no classification analogous to litters in the example of §4.15. The 5 d.f. between doses can