THE COMPARISON OF DIFFERENT SCALES OF MEASUREMENT FOR EXPERIMENTAL RESULTS[1,2]

By W. G. Cochran

Iowa State College

1. Introduction. In some fields of research, the development of a satisfactory method for measuring the effects of experimental treatments constitutes a difficult problem. The estimation of the vitamin content of preparations of foods furnishes a good example; for most of the vitamins several years of work were required to construct a reliable method of assay. In other cases, where the ideal method for measuring treatment responses is costly or troublesome, a search may be made for a more convenient substitute. Thus in pasture or forage-crop experiments the species composition of a plot may be estimated by eye inspection as a substitute for a complete botanical separation. As a third example we may quote experiments in cookery, where the flavor and quality of the dishes are subject to the whims of human taste. Frequently a panel of judges is employed, each of whom scores the dishes independently. It is not easy to determine how the panel should be chosen, nor how representative its verdicts are of consumer preferences in general.

When such problems are investigated, experiments may be carried out specifically for the purpose of comparing two or more methods or *scales* of measurement. Where the process of measurement affects only the final stages of the experiment, as in the last two examples quoted above, all that is necessary is to score the *same* experiment by the various scales under consideration. In comparing two different methods of assaying vitamins, on the other hand, independent experiments are frequently required, the only common feature being that the same set of treatments is tested in both experiments.

In the interpretation of the results of such experiments, two types of comparison are of general interest. One concerns the *relations* between the scales. It may be summed up rather loosely in the question: Are the effects of the treatments the same in all scales? For a more exact formulation, consider the case of two scales, which is probably the most frequent in practice. Let ξ_{1t}, ξ_{2t} be the true means of the tth treatment as measured on the two scales. We may wish to examine the following hypotheses:

(i) *Scales equivalent:*

$$\xi_{1t} = \xi_{2t}, \qquad \text{(all } t\text{)}; \tag{1}$$

(ii) *Scales equivalent, apart from a constant difference:*

$$\xi_{1t} = \xi_{2t} + \epsilon, \qquad \text{(all } t\text{)}; \tag{2}$$

[1] Paper presented at a meeting of the Institute of Mathematical Statistics, Washington, D. C., June 18, 1943.

[2] Journal Paper No. J-1136 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project 514.

(iii) *Scales linearly related:*

$$\alpha\xi_{1t} + \beta\xi_{2t} = \gamma, \qquad \text{(all } t\text{)}; \tag{3}$$

(iv) *Relation monotonic, but not linear:*

$$\xi_{1t} = f(\xi_{2t}, \alpha, \beta, \cdots), \qquad \text{(all } t\text{)}; \tag{4}$$

where the function is strictly monotonic.

In this case the two scales are mutually consistent in that they place any set of treatments in the same order. The ratio of a treatment difference in one scale to the corresponding difference in the other scale is, however, not constant.

(v) *Relation not monotonic:* Here the scales do not place the treatments in the same order and consequently are not satisfactory substitutes for each other.

The second question concerns the relative *accuracy* or *sensitivity* of the two scales. For practical purposes this question may be put as follows: how many replications are required with the second scale to attain the accuracy given by r replications with the first scale? It is clear that the answer depends both on the experimental errors associated with the scales and on the magnitudes of the treatment effects in the two scales. For example, Coward [1] reports that in the assay of vitamin D, male rats give a higher experimental error than females, yet provide a more accurate assay because they are more responsive. The relative accuracy may be different in different parts of the two scales. This is likely to happen whenever the relation between the scales is of type (iv) above.

This paper gives a preliminary discussion of some of the simpler questions raised above, to which recent work in multivariate analysis is applicable. A complete solution for small sample work appears to demand considerable further development in the distribution theory of multivariate analysis.

The discussion is confined to the case in which all scales measure the same experiment. The case where each scale requires a separate experiment may be expected to be somewhat simpler, but cannot conveniently be treated as a special case of the procedure for a single experiment.

2. Assumptions. Let $x_1, x_2, \cdots x_p$ denote measurements on the p scales and let n_1 and n_2 be the numbers of degrees of freedom for treatments and error respectively. The experimental data furnish a joint analysis of variance and covariance of the p variates as follows:

(5)

	d.f.	*Sum of squares or products*
Mean	1	m_{ij}
Treatments	n_1	a_{ij}
Error	n_2	b_{ij}

It will be assumed that $x_1, \cdots, x_p$ follow a multivariate normal distribution, and that for any pair of variates x_i, x_j the error mean covariance σ_{ij} is constant throughout the experiment (though it may vary as i and j vary). Thus the

quantities b_{ij} follow the standard joint distribution, Wishart [16], of sums of squares and products while the quantities m_{ij} and a_{ij} follow the corresponding non-central distributions and the three sets of distributions are independent.

3. Tests for equivalence. If there are only two scales, a test for equivalence is obtained from elementary techniques. An analysis of variance similar to (5) is computed on the *differences* between the two scales for every observation. If equations (1) hold in the population, the sums of squares for the Mean, Treatments and Error are distributed independently as $\chi^2(\sigma_{11} + \sigma_{22} - 2\sigma_{12})$. The pooled mean square for the Mean and Treatments may therefore be compared with the Error mean square in a variance-ratio test, the degrees of freedom being $(n_1 + 1)$ and n_2. If the scales are equivalent apart from a constant difference, the same result is valid for Treatments and Error, while the mean square for the Mean is proportional to a non-central χ^2. Thus separate z- or F-tests on the Mean and Treatments assist in distinguishing between hypotheses (1) and (2).

4. More than two scales. Let ξ_{it} be the true mean of the tth treatment as measured on the ith scale. The first two hypotheses may now be written respectively:

$$\xi_{it} = \xi_t \tag{1'}$$

$$\xi_{it} = \xi_t + \epsilon_i \tag{2'}$$

for $i = 1, 2, \cdots, p$. The quantities ϵ_i, whose sum may be assumed zero, measure the constant differences among the scales.

If the interactions of all components with Scales are computed, the analysis of variance extends formally, with the following separation of degrees of freedom:

(6)

	d.f.
Mean × Scales	$(p - 1)$
Treatments × Scales	$n_1(p - 1)$
Error	$n_2(p - 1)$

The three lines in the analysis play the same roles as before in relation to hypotheses (1') and (2'). When $p > 2$, however, it may be shown that the three sums of squares are not distributed as multiples of χ^2 unless (i) all scales have the same error variance and (ii) every pair of scales has the same correlation coefficient. Where these conditions are reasonably well satisfied, as happens possibly when experienced judges employ a similar scoring system, the above analysis supplies approximate tests. But with scales which differ widely in their experimental errors or in their degrees of interrcorrelation, the validity of variance-ratio tests is open to more serious question.

In order to obtain an exact test, we may note that hypothesis (1') is closely related to the Wilks-Lawley hypothesis (Wilks [15], Lawley [9], Hsu [7]) that the means of k populations are all equal. If each treatment denotes a separate population, the Wilks-Lawley hypothesis states that

$$\xi_{it} = \xi_i \qquad (t = 1, 2, \cdots, n_1 + 1). \tag{7}$$

Since this differs from (1′) only in the interchange of the letters i and t, it is clear that the two hypotheses may be subjected to the same kind of test.

For the details of the procedure we first divide the $(p-1)$ comparisons among scales into $(p-1)$ single comparisons by the introduction of a set of variates y_i, $(i = 1, 2, \cdots, p-1)$.

$$y_i = \sum_{j=1}^{p} \lambda_{ij} x_j . \tag{8}$$

Any set of y's may be chosen, provided that they are linearly independent and that

$$\sum_{j=1}^{p} \lambda_{ij} = 0, \qquad (i = 1, 2, \cdots (p-1)). \tag{9}$$

Thus with three scales we might use $y_1 = x_1 - x_2$, $y_2 = x_1 - x_3$ or $y_1 = 2x_1 - x_2 - x_3$, $y_2 = x_2 - x_3$.

The next step is to compute an analysis of variance and covariance of the y variates, as follows:

(10)

	d.f.	*Sum of squares or products*
Mean	1	m'_{ij}
Treatments	n_1	a'_{ij}
Error	n_2	b'_{ij}

If hypothesis (1′) holds, it follows from (9) that the three sets of quantities m'_{ij}, a'_{ij} and b'_{ij} all follow the standard joint distribution for sums of squares and products. Hence Wilks' test (Wilks [15], Pearson and Wilks [11], Hsu [7]), for the equality of the means of k populations may be applied. For a single test of hypothesis (1′) we may use

$$W = \frac{|b'_{ij}|}{|b'_{ij} + m'_{ij} + a'_{ij}|} . \tag{11}$$

As before, if W is significant we may test whether the deviation is due to constant differences or to other types of difference among the scales by calculating

$$W_m = \frac{|b'_{ij}|}{|b'_{ij} + m'_{ij}|} , \tag{12}$$

and

$$W_t = \frac{|b'_{ij}|}{|b'_{ij} + a'_{ij}|} . \tag{13}$$

The flexibility of analysis of variance tests is not sacrificed; in particular we may test any desired subgroup of the treatments or of the scales. When there are only two scales the tests reduce to those given in section 3.

The tests are invariant under homogeneous linear transformations of the y's

which explains why the form of the subdivision of the scale comparisons is immaterial. In fact for purposes of computation it is not necessary to introduce the y's. By taking a simple transformation and expressing a'_{ij} in terms of a_{ij}, etc., we may express W directly in terms of the x's, as follows:

$$W = \frac{\sum_{ij} B_{ij}}{\sum_{ij} (B + M + A)_{ij}}, \tag{14}$$

where B_{ij}, $(B + M + A)_{ij}$ are respectively the co-factors of the matrices (b_{ij}), $(b_{ij} + m_{ij} + a_{ij})$. Analogous expressions hold for W_m and W_t. In practice it will often be preferable to compute the y's in order that particular comparisons among the scale variates may be examined in detail.

The form of the frequency distribution has been worked out by Wilks [15]. For small values of n_1 and p, the test of significance can be referred to the recent tables of the significance levels of the incomplete Beta-function, Thompson [13], or to variance-ratio tables. Such cases are listed below, from Wilks [15] and Hsu [7]. In our notation, ν_1 is taken as $(n_1 + 1)$ in equation (11), as 1 in equation (12) and as n_1 in equation (13).

$\underline{p = 3, \nu_1 > 1}$: $f(W) \propto W^{\frac{1}{2}(n_2-3)}(1 - W^{\frac{1}{2}})^{\nu_1-1}$

$$: F\{2\nu_1, 2(n_2 - 1)\} = \frac{(n_2 - 1)(1 - W^{\frac{1}{2}})}{\nu_1 W^{\frac{1}{2}}},$$

$\underline{\nu_1 = 1}$: $f(W) \propto W^{\frac{1}{2}(n_2-p)}(1 - W)^{\frac{1}{2}(p-3)}$

$$: F\{p - 1, n_2 - p\} = \frac{(n_2 - p)(1 - W)}{(p - 1)W}.$$

This distribution applies to all tests made on the Mean, equation (12), and all cases where a single degree of freedom is isolated from the treatment comparisons.

$\underline{\nu_1 = 2}$: $f(W) \propto W^{\frac{1}{2}(n_2-p)}(1 - W^{\frac{1}{2}})^{p-2}$

$$: F\{2(p - 1), 2(n_2 - p + 2)\} = \frac{(n_2 - p + 2)(1 - W^{\frac{1}{2}})}{(p - 1)W^{\frac{1}{2}}}.$$

A tabulation of the distributions for four and five scales would be useful. Hsu [7] has shown that as n_2 becomes large, the distribution of $-n_2 \log W$ tends to that of χ^2 with $\nu_1(p - 1)$ degrees of freedom. In general, this approximation does not agree very well with the exact distributions above unless n_2 exceeds 60.

5. Interpretation as a problem in canonical correlations. As an introduction to the methods that will be used in testing the hypothesis of linearity, we may note that hypotheses (1′) and (2′) can be described in terms of canonical correlations. Fisher [5] has pointed out that the roots θ of the equation

$$| a_{ij} - \theta(a_{ij} + b_{ij}) | = 0, \tag{15}$$

are the squares of the *sample* canonical correlations between the x-variates and a set of n_1 dummy variates which represent the n_1 degrees of freedom among treatments. In order to obtain the corresponding equation for the *population* correlations, we may suppose that n_1 and p remain constant while the number of replicates r' and consequently n_2 increase without limit. After the removal of a common factor r', equation (15) becomes

$$| \psi_{ij} - \rho^2(\psi_{ij} + \nu\sigma_{ij}) | = 0, \tag{16}$$

where

$$\psi_{ij} = \sum_{t=1}^{n_1+1} (\xi_{it} - \bar{\xi}_i)(\xi_{jt} - \bar{\xi}_j). \tag{17}$$

The value of the coefficient ν depends on the type of experimental design. For a randomized block layout, $\nu = n_1$ and for a simple group comparison $\nu = (n_1 + 1)$.

Now if hypothesis (2′) is true, i.e., $\xi_{it} = \xi_t + \epsilon_i$, it follows that ψ_{ij} is independent of i and j. In this event equation (16) has $(p - 1)$ roots ρ^2 which are identically zero. The remaining root corresponds to the best discriminant function, Fisher [5], and does not vanish unless the treatments have no effects on any of the x-variates.

Let $\Sigma\beta_i x_i$ be a population canonical variate for the scale variables. The coefficients β_i satisfy the equations

$$\sum_j \beta_j\{\psi_{ij} - \rho^2(\psi_{ij} + \nu\sigma_{ij})\} = 0. \qquad i = 1, \cdots p. \tag{18}$$

For a zero root $\rho^2 = 0$ we have $\psi_{ij} =$ constant. Hence if a zero root is substituted, equation (18) degenerates into

$$\beta_1 + \beta_2 + \cdots + \beta_p = 0. \tag{19}$$

To summarize, hypothesis (2′) specifies that (i) $(p - 1)$ of the population canonical correlations vanish and (ii) any variate $\Sigma\beta_i x_i$ is a canonical scale variate corresponding to a zero root, provided that equation (19) is satisfied. Analogous results hold for hypothesis (1′); in this case we replace the Treatments line of the analysis of variance by the (Treatments + Mean) line.

6. Test for linear relationship—two scales. We may assume $n_1 \geq 2$; otherwise no test of linearity is possible. If the values of α, β and γ in equations (3) are known, the problem can be reduced to that of testing hypothesis (1) or (2). Since this case is unlikely to be encountered frequently in practice, further details are omitted.

When α, β and γ are unknown, we may theoretically replace the variates x_1 and x_2 by $v_1 = \alpha x_1 + \beta x_2$ and $v_2 = \mu_1 x_1 + \mu_2 x_2$, where μ_1 and μ_2 are chosen so that v_1 and v_2 are independently distributed. If hypothesis (3) holds, it follows from (17) that in terms of the v's, $\psi_{11} = \psi_{12} = 0$. Since in addition $\sigma_{12} = 0$, the two roots of equation (16) are

$$\rho^2 = 0 \quad \text{and} \quad \rho^2 = \psi_{22}/(\psi_{22} + \nu\sigma_{22}). \tag{20}$$

Thus hypothesis (3) implies that one of the population canonical correlations vanishes. Unlike the previous case, however, we cannot construct the corresponding canonical variate, which requires knowledge of α and β.

The selection of a sample test criterion opens up some difficulties. Pending further elucidation of the problem, the natural choice seems to be the square r_2^2 of the lower sample canonical correlation, or the equivalent quantity $h_2 = r_2^2/(1 - r_2^2)$, where h_2 is the lower root of the equation:

$$| a_{ij} - hb_{ij} | = 0. \tag{21}$$

It appears likely, however, that r_2^2 and h_2 are not sufficient estimates of the corresponding population parameters.

When n_2 is large, Hsu [8] has shown that the distribution of $n_2 h_2$ tends to that of χ^2 with $(n_1 - 1)$ degrees of freedom. A considerable advance towards the small-sample distribution is obtainable from Madow [10], who developed an expression for the exact distribution of r_1^2 and r_2^2 when one of the population correlations is different from zero. In our notation this result, which is an important generalization of the distribution found by Fisher [5] and Girshick [6] may be written as follows:

$$\frac{(n_1 + n_2 - 2)!}{4\pi(n_1 - 2)!\,(n_2 - 2)!}\,(r_1^2 r_2^2)^{\frac{n_1-3}{2}}\,\{(1 - r_1^2)(1 - r_2^2)\}^{\frac{n_2-3}{2}}\,(r_1^2 - r_2^2)\,dr_1^2\,dr_2^2$$

$$\times\,(1 - \rho_1^2)^{\frac{1}{2}(n_1+n_2)} \int_{r_2^2}^{r_1^2} \frac{F\left(\frac{n_1 + n_2}{2}, \frac{n_1 + n_2}{2}, \frac{n_1}{2}, \rho_1^2 y\right) dy}{\sqrt{(r_1^2 - y)(y - r_2^2)}}, \tag{22}$$

where ρ_1 is the non-vanishing population correlation. It is evident from the form of (22) that the distribution of r_2^2 or h_2 involves ρ_1. The conditional distribution of h_2/h_1 may be relatively insensitive to changes in ρ_1, though even this distribution does not seem entirely independent of ρ_1.

When ρ_1 is unity, the small-sample distribution of h_2 is that of the ratio of two independent sums of squares, i.e., $h_2 = (n_1 - 1)e^{2z}/n_2$, with $(n_1 - 1)$ and n_2 degrees of freedom. This result is a particular case of a more general result proved in section 8. From (20) it is seen that ρ_1 is close to unity when ψ_{22} is large relative to σ_{22}, i.e., when the real differences among the treatments are large relative to the experimental errors. In the absence of a usable exact solution, the F-distribution may be a better approximation than the large-sample distribution of h_2 for data where r_1 is found to be close to unity, though proof of this statement is not yet available.

If it is desired to test hypothesis (3) with the additional assumption that $\gamma = 0$, we replace a_{ij} by $(a_{ij} + m_{ij})$ in equation (21) for h_2, and n_1 by $(n_1 + 1)$ in the distribution theory.

7. Connection with the method of least squares. The previous approach has an interesting connection with the method of least squares. We are required to test the linearity of relationship between $(n_1 + 1)$ pairs of means $(\bar{x}_{1t}, \bar{x}_{2t})$.

Both variates are subject to error and the errors are correlated; with r' replications the population variances and covariance of these means are σ_{11}/r', σ_{22}/r' and σ_{12}/r'. For these unknown quantities we have sample estimates b_{11}/n_2r', b_{22}/n_2r' and b_{12}/n_2r' respectively, derived from the Error line of the analysis of variance.

The procedure suggested by the method of least squares is to estimate the parameters of the line and use the deviations of the points $(\bar{x}_{1t}, \bar{x}_{2t})$ from the line for a test of linearity. If the population variances were known, the unknown quantities α, β, γ and ξ_{it} would be estimated by minimizing the quadratic form:

$$(23)\quad \sigma^{11}\sum_{t=1}^{n_1+1} r'(\bar{x}_{1t}-\xi_{1t})^2 + 2\sigma^{12}\sum_{t=1}^{n_1+1} r'(\bar{x}_{1t}-\xi_{1t})(\bar{x}_{2t}-\xi_{2t}) + \sigma^{22}\sum_{t=1}^{n_1+1} r'(\bar{x}_{2t}-\xi_{2t})^2,$$

subject to the linear relations (3). Here (σ^{ij}) is the matrix inverse to σ_{ij}. On substitution of the estimates, expression (23), which is positive definite, would serve as a "sum of squares" of deviations from the line and therefore as a test criterion. This criterion is of course a direct generalization of the weighted sum of squares which is used when the errors are independent.

Van Uven [14] gave an elegant method by which the sum of squares of deviations can be found directly, before solving for any of the unknown quantities In our notation he showed that the sum of squares of deviations is the smaller root H_2 of the equation

$$(24)\qquad |\, a_{ij} - H\sigma_{ij} \,| = 0,$$

where a_{ij} is as before the treatments sum of squares or products.

Suppose that in default of knowledge of the σ_{ij} we derive the weights from the sample estimates b_{ij}/n_2; i.e., we minimize (23) with b^{ij} in place of σ^{ij}, where $(b^{ij}) = (b_{ij}/n_2)^{-1}$. In this case the method of Van Uven shows that the sum of squares of deviations from the best-fitting line is the smaller root H_2' of the equation

$$(25)\qquad \left| a_{ij} - \frac{H'}{n_2} b_{ij} \right| = 0.$$

Comparing (25) with (21) we find $H_2' = n_2h_2$. Consequently the least squares approach, with sample weights substituted in (23) for the unknown true weights, leads to h_2 as a test criterion. Further, Hsu's [8] proof that the distribution of n_2h_2 tends to χ^2 with $(n_1 - 1)$ degrees of freedom establishes for this case the standard least-squares result for the distribution of the residual sum of squares: —namely that when the population weights are known, the residual sum of squares is distributed as χ^2, with degrees of freedom equal to the number of points, $2(n_1 + 1)$, *minus* the number of independent unknowns, $(n_1 + 3)$. By a transformation of the x-variates to independent variables, this result can be obtained alternatively from a theorem by Deming [2].

8. Test for linear relationship—more than two scales. The extension of hypothesis (3) to the case of p scales can be expressed by means of the equations

$$\alpha_i\xi_{1t} + \beta_i\xi_{it} = \gamma_i : \qquad (i = 2, \cdots p)(t = 1, \cdots n_1 + 1). \tag{3'}$$

The equations, $(p - 1)(n_1 + 1)$ in number, postulate a linear relation between x_1 and every other variate and consequently imply a linear relation between any pair of variates x_i and x_j.

Consider the variates $v_i = \alpha_i x_{1t} + \beta_i x_{it}$, $(i = 2, \cdots p)$. For v_1 we choose the linear function of the x's which is independent of $v_2, \cdots v_p$. Thus in equation (16) for the population canonical correlations we have $\psi_{ij} = 0$, $(i, j, \geq 2)$ and $\sigma_{1j} = 0$, $(j > 1)$. It follows that all roots of equation (16) are zero except one, the non-vanishing root being $\rho^2 = \psi_{11}/(\psi_{11} + \nu\sigma_{11})$. If each treatment denotes a separate population, hypothesis (3') is therefore identical with Fisher's hypothesis [4], that the populations are *collinear*.

As a test criterion for this hypothesis Fisher has suggested the sum of the roots of equation (21), excluding the highest root, i.e., $V' = \Sigma h_i = \Sigma r_i^2/(1 - r_i^2)$. If $n_1 \geq p$ the sum extends over $(p - 1)$ roots, while if $n_1 < p$ the sum extends over $(n_1 - 1)$ roots. For computational purposes it may be more expeditious to form this sum by subtraction. Hsu [7] has pointed out that the sum of all roots is given by $V = \sum_{ij} b^{ij}a_{ij}$, which is obtained readily when the inverse of (b_{ij}) has been calculated. The largest root of (21) is then found and subtracted from V.

Fisher [4] also suggested that when equations (3') hold, the distribution of V' is approximately that of χ^2 with $(p - 1)(n_1 - 1)$ degrees of freedom. This result has been confirmed by Hsu [8] as the limiting form of the V' distribution when n_2 tends to infinity. As in the case of two scales, the small-sample distribution is as yet unknown; it presumably contains ρ_1, the non-vanishing correlation, as a nuisance parameter.

Some progress towards the small-sample distribution can be made without difficulty in the case where $\rho_1 = 1$. For then v_1 must have a zero Error sum of squares in every sample from the population, i.e., v_1 is constant within any given treatment. Consequently (i) $b_{1i} = 0$ for $i = 1, \cdots p$, and (ii) a_{1j}^2/a_{11} is a single degree of freedom from the Treatments sum of squares of v_j. On account of conditions (i), equation (21) reduces to

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{12} & a_{22} - hb_{22} & \cdots & a_{2p} - hb_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1p} & a_{2p} - hb_{2p} & \cdots & a_{pp} - hb_{pp} \end{vmatrix} = 0. \tag{26}$$

Subtract a_{1i}/a_{11} times the first row from the ith row, for $i = 2, \cdots p$. We see that one root is infinite; the rest are the roots of the equation

$$|a''_{ij} - hb_{ij}| = 0, \qquad i, j = 2, \cdots p, \tag{27}$$

where $a''_{ij} = a_{ij} - a_{1i}a_{1j}/a_{11}$.

If hypothesis (3′) holds, the quantities a''_{ij} follow the Wishart distribution [16] with $(n_1 - 1)$ degrees of freedom. Hence the joint distribution of $h_2, \cdots h_p$ or h_{n_1}, is that which is obtained when all the population canonical correlations vanish, with $(n_1 - 1)$ in place of n_1. For $n_1 \geq p$, the distribution function (apart from the constant term) is:

$$\prod_{i=2}^{p}\left[h_i^{\frac{1}{2}(n_1-p-1)}(1+h_i)^{-\frac{1}{2}(n_1+n_2-1)}\left\{\prod_{j=i+1}^{p}(h_i-h_j)\right\}\right]. \tag{28}$$

For two scales, $(p = 2)$, we reach the result mentioned in section 6, that $V' = h^2$ is distributed as $(n_1 - 1)e^{2s}/n_2$. This result can also be obtained directly from (27). When $p = 3$, the distribution of V' is obtainable from a result by Hsu [7].

9. Measures of relative sensitivity. We propose to discuss briefly the estimation of the relative sensitivity of two scales and to indicate the types of distribution that are involved. If there are only two treatments, t, t', an appropriate definition of the true sensitivity of the ith scale is

$$\frac{(\xi_{it'} - \xi_{it})^2}{2\sigma_{ii}}, \tag{29}$$

or some simple function of this quantity. In justification, we may observe that for a fixed number of replicates, the power function of the t-test in the ith scale depends entirely on this quantity. An unbiased sample estimate is

$$\frac{(n_2 - 2)(\bar{x}_{it'} - \bar{x}_{it})^2}{2b_{ii}} - \frac{1}{r'}, \tag{30}$$

where r' is the number of replicates. Since (30) involves a non-central variance ratio, confidence limits for the true sensitivity can be found from Fisher's Type C distribution, Fisher [3].

It follows from (3) and (29) that if two scales are linearly related (including the case of equivalence) their relative sensitivity is constant for all treatment comparisons. For scale 1 relative to scale 2 the sensitivity is measured by $\beta^2\sigma_{22}/\alpha^2\sigma_{11}$.

If the scales are equivalent, apart possibly from a constant difference, this quantity reduces to $\varphi = \sigma_{22}/\sigma_{11}$, for which $F = b_{22}/b_{11}$ serves as a sample estimate. A test of significance of the sample ratio and confidence limits for the true ratio may be obtained from Pitman [12], who showed that

$$\left(\frac{F}{\varphi} - 1\right) \Big/ \sqrt{\left(\frac{F}{\varphi} + 1\right)^2 - \frac{4r_{12}^2 F}{\varphi}}, \tag{31}$$

follows the distribution of a sample correlation coefficient from $(n_2 + 1)$ pairs of observations. In (31), $r_{12}^2 = b_{12}^2/b_{11}b_{22}$. The same procedure may be used whenever α and β are known.

When α and β are unknown, a sample estimate of the relative sensitivity is b^2b_{22}/a^2b_{11}, where $(ax_1 + bx_2)$ is the discriminant function which corresponds to

the lower root of equation (21). We have not been able to reach the distribution of this estimate. Confidence limits for the relative sensitivity can, however, be obtained when n_2 is sufficiently large so that σ_{11} and σ_{22} may be assumed known. For in that case the problem reduces to that of finding confidence limits for β^2/α^2. Now if α, β are the true coefficients, the quantity

$$\frac{\alpha^2 a_{11} + 2\alpha\beta a_{12} + \beta^2 a_{22}}{\alpha^2 b_{11} + 2\alpha\beta b_{12} + \beta^2 b_{22}}, \tag{32}$$

follows the $n_1 e^{2z}/n_2$ distribution. Any proposed values of α and β which make (32) significant are rejected by the evidence of the sample. By equating (32) to the desired significance level of $n_1 e^{2z}/n_2$, we get a quadratic equation for the two limits of β/α. The limits will not be narrow unless the treatment effects are large.

If the relation between the scales is non-linear, and the assumption of a constant error variance throughout an individual scale is valid, the relative sensitivity differs for different treatment comparisons. Even in this event estimates of relative sensitivity may be of interest. Attention might be restricted to a single degree of freedom from the treatment comparisons, in which case the definition for two treatments could be applied.

Alternatively an estimate might be wanted of the *average* relative sensitivity over all treatment comparisons. For a given number of replicates, the power function of the variance-ratio test of the treatment effects in the ith scale depends only on the quantity

$$\frac{\sum_t (\xi_{it} - \bar{\xi}_i)^2}{\sigma_{ii}}. \tag{33}$$

Consequently this quantity, which is an extension of (29), might be chosen as a measure of average sensitivity. The corresponding generalization of the unbiased sample estimate (20) is

$$\frac{(n_2 - 2)a_{ii}}{n_1 r' b_{ii}} - \frac{1}{r'}. \tag{34}$$

Since the quantity a_{ii}/b_{ii} is a multiple of a non-central variance ratio, the comparison of two scales involves a test of significance of the hypothesis that two non-central variance ratios are equal.

10. Summary. This paper discusses the analysis of data obtained when the results of a replicated experiment are measured on several different scales which we wish to compare. Recent work in multivariate analysis provides tests of the hypothesis that the treatment effects are the same in all scales, and of the hypothesis that the scales are linearly related. When the number of Error degrees of freedom is large, the significance levels of these tests are obtainable from the standard tables. For small sample tests, further investigation and

tabulation of certain distributions will be needed, particularly that of the sample canonical correlations when one population correlation differs from zero.

A brief discussion is given of methods for comparing the relative sensitivity of two scales.

REFERENCES

[1] K. H. COWARD, *The Biological Standardization of the Vitamins*, 1937.
[2] W. E. DEMING, "The chi-test and curve-fitting," *Jour. Amer. Stat. Assn.*, Vol. 29 (1934), pp. 372–382.
[3] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc. A*, Vol. 121 (1928), pp. 654–673.
[4] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol. 8 (1938), pp. 376–386.
[5] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 238–249.
[6] M. A. GIRSHICK, "On the sampling theory of the roots of determinental equations," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 203–224.
[7] P. L. HSU, "On generalized analysis of variance (I)," *Biometrika*, Vol. 31 (1940), pp. 221–237.
[8] P. L. HSÜ, "The problem of rank and the limiting distribution of Fisher's test function," *Annals of Eugenics*, Vol. 11 (1941), pp. 39–41.
[9] D. N. LAWLEY, "A generalization of Fisher's z-test," *Biometrika*, Vol. 30 (1938), pp. 180–187.
[10] W. G. MADOW, "Contributions to the theory of multivariate statistical analysis," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), p. 490.
[11] E. S. PEARSON and S. S. WILKS, "Methods of statistical analysis appropriate for k samples of two variables," *Biometrika*, Vol. 25, (1933), pp. 353–378.
[12] E. J. G. PITMAN, "A note on normal correlation," *Biometrika*, Vol. 31 (1939), pp. 9–12.
[13] C. M. THOMPSON, "Tables of percentage points of the incomplete Beta-function," *Biometrika*, Vol. 32 (1942), pp. 151–181.
[14] M. J. VAN UVEN, "Adjustment of N points (in n-dimensional space) to the best linear $(n-1)$ dimensional space," *Proc. Koninklizke Akad. van Wetenschappen te Amsterdam*, Vol. 33 (1930), pp. 143–157.
[15] S. S. WILKS, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.
[16] J. WISHART, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, Vol. 20A (1928), pp. 32–52.