SOME STATISTICAL METHODS FOR COMPARISON OF GROWTH CURVES

C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta, India

INTRODUCTION

In experiments involving study of growth the observations on a growing organism can sometimes be obtained continuously in time as a curve but generally, and more conveniently, at a finite number of specified time points. Two problems of interest which can be studied from such data are as follows. One is to construct a simple stochastic model characterising the growth of an individual organism during a certain period of time. Another is to compare the characteristics of growth under different conditions such as diet, environment, etc. If we succeed in obtaining a simple growth model, the second problem leads to the comparison of the models applicable to different situations. But this is not absolutely necessary for comparison of growth curves for we might compare various physically definable and meaningful aspects of growth such as, total growth in a period, average growth rate, changes in growth rate although they may not completely explain the growth processes.

An early example of comparison of growth curves is due to Wishart [1938]. To each individual growth curve classified by litter, sex, and treatment, a second-degree polynomial in time was fitted by the least squares method. The coefficients of linear and quadratic terms were taken to represent salient features of growth and the fifteen or so observations on an individual growth curve were replaced by these two coefficients. The analysis then consisted in comparing the mean values of these coefficients under different experimental conditions. The analysis was justified because a large portion of the differences in growth curves was concentrated in the linear growth rate and to a lesser extent in the differential rate of growth measured by the coefficient of the quadratic term.

Comparison of mean values at fifteen time points instead of these two aspects of growth would be a less efficient procedure although valid multivariate tests exist for such comparisons. The success then consists in replacing the various observations on growth by a few summary figures which lead to most efficient comparisons between groups. This is essential if comparisons have to be made on the basis of small samples and variations between individual growth curves are uncontrollably large. In fact, in such situations, effort should be made to reduce the data to the lowest possible number of dimensions without sacrificing the essential information. The purpose of this paper is to explore such possibilities and to develop the necessary tests of significance.

1. A SIMPLE AND EXACT ANALYSIS FOR COMPARISON OF GROWTH CURVES.

1.1. *Comparison of rates of growth*

As stated above, the main problem, especially in small samples, is to obtain an adequate representation of a growth curve with the minimum possible number of factors on the basis of which significant differences could be established between differently treated groups of individuals. The emphasis at this stage is not on obtaining a model adequately describing the growth of an individual but on examining whether differences exist between groups of growth curves. It should, however, be noted that, with small samples, it will be impossible to discriminate among a large number of widely varying models.

Let us replace the observations on growth at different time points by the initial value and successive differences giving the gain in growth in different periods. Symbolically they may be represented by

$$y_0 \, , y_1 \, , y_2 \, , \tag{1}$$

If the growth rate is uniform during the period under study, it is possible to replace the series (1) by the initial value and an estimate of the growth rate. With these two variables, comparisons between groups can be carried out. Rate is rarely uniform and in general growth is a complicated function of time. In controlled experiments, it may be a monotonic decreasing function of time during the period of growth. If, however, time can be transformed by a function $\tau = G(t)$ in such a way that the growth rate is uniform with respect to the chosen time *metameter*, then an adequate representation is available in terms of the initial value and the redefined *uniform rate*.

Let us represent the length of the interval between the $(i - 1)$-th and i-th time points on the transformed time axis by g_i. The increase

y_i then corresponds to the time period g_i so that an estimated rate of growth with respect to τ is*

$$b = \sum y_i g_i / \sum g_i^2$$

and is proportional to $\sum y_i g_i$. The set of observations representing an individual's growth can then be replaced by y_0 and b. If the problem involves comparison of growth under different conditions, we need to test whether the mean value of b is the same in all groups by analysis of variance with respect to the single variable b, eliminating the initial value y_0 by analysis of covariance, if necessary. Thus there appears to be no difficulty when the g_i are known.

Fortunately the same test seems to be valid even if the g_i are estimated from the data themselves in a certain way and this is obviously better than depending on a priori values of the g_i . The estimate of g_i is taken to be the grand mean** of y_i , the gain in the i-th interval, for all individuals included in the sample. The analysis of variance test is exact for such a choice of g_i under the assumption of normality of the distribution of y_i . The proof is immediate if we observe that analysis of variance depends on the differences in the averages, while estimates of g_i , being based on totals, are distributed independently of these differences. We shall illustrate this method using the observations on the growth of rats under three different conditions given in a paper by Box [1950].

The totals of 27 observations for each week provide estimates of g_1 , g_2 , g_3 , and g_4 .

$$\theta_1^* = \sum y_1 = 603, \qquad \theta_3 = \sum y_3 = 570,$$

$$\theta_2 = \sum y_2 = 673, \qquad \theta_4 = \sum y_4 = 674.$$

From these, the value of $b = \sum y_i \theta_i$ is computed for each rat and given in Table 1 (after dividing by 1000 arbitrarily to reduce the scale) along with the observed values of gains in weight in the successive weeks.

The analysis of variance and covariance for b and y_0 is given in Table 2.

---

*This resembles the regression estimate. The average of (y_i/g_i) is another estimate and so also the simple average of y_i . The appropriateness of the formula employed depends on the assumptions made on the variances of y_i. The method of testing developed here is valid for all these types of estimates.

**The method of estimation was proposed by G. Rasch of Denmark during a course of lectures on growth curves which he gave in India in 1951. This can be shown to be the least squares estimate of the time metameter corresponding to the observed values under the assumption that the y_i are uncorrelated and have the same variance.

TABLE 1

INITIAL WEIGHT AND WEEKLY GAINS IN WEIGHTS OF RATS UNDER THREE DIFFERENT TREATMENTS

No.	y_0	y_1	y_2	y_3	y_4	b
	Group 1, Control					
1	57	29	28	25	33	72.82
2	60	33	30	23	31	74.09
3	52	25	34	33	41	84.40
4	49	18	33	29	35	73.18
5	56	25	23	17	30	60.46
6	46	24	32	29	22	67.37
7	51	20	23	16	31	57.55
8	63	28	21	18	24	57.45
9	49	18	23	22	28	57.74
10	57	25	28	29	30	70.67
	Group 2, Thyroxin					
1	59	26	36	35	35	83.45
2	54	17	19	20	28	53.31
3	56	19	33	43	38	83.79
4	59	26	31	32	29	74.33
5	57	15	25	23	24	55.16
6	52	21	24	19	24	55.82
7	52	18	35	33	33	75.46
	Group 3, Thiouracil					
1	61	25	23	11	9	42.89
2	59	21	21	10	11	39.91
3	53	26	21	6	27	51.43
4	59	29	12	11	11	39.25
5	51	24	26	22	17	55.97
6	51	24	17	8	19	43.28
7	56	22	17	8	5	32.64
8	58	11	24	21	24	50.93
9	46	15	17	12	17	38.78
10	53	19	17	15	18	43.58

y_0 = initial weight
y_1 = gain in 1st week
y_2 = gain in 2nd week
y_3 = gain in 3rd week
y_4 = gain in 4th week
$b = \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \beta_4 y_4$

TABLE 2

Analysis of Variance and Covariance for b and y_0

Source	d.f.	S_{bb}	S_{by_0}	$S_{y_0y_0}$	S_{bb} corrected for y_0	d.f.	Mean sq.	F
Between groups	2	3691.87	− 0.39	10.19	3691.97	2	1845.99	18.5
Within	24	2297.01	35.98	517.81	2294.51	23	99.76	
Total	26	5988.88	35.59	528.00	5986.48	25		

The variance ratio is significant showing that growth rates are different. An examination of the mean values of the regression coefficients for the three groups,

$$\bar{b}_1 = 67.575, \qquad \bar{b}_2 = 68.758, \qquad \bar{b}_3 = 43.865,$$

shows that the differences are mainly due to the smaller rate for the third group. The use of covariance analysis was not worthwhile because of an extremely poor or no correlation between b and y_0 . It may also be noted that the sum of squares between groups is practically unchanged when corrected for y_0 This is probably due to some sort of balancing with respect to average initial weight in assigning the rats to the three groups.

The following comments about the test proposed above are worth noting:

(i) It provides a valid test of the null hypothesis that the average growth curve is the same under all treatment conditions irrespective of any assumptions on the nature of the growth curve. If the average growth curve can be represented by a straight-line trend for each group by a suitable choice of a common time metameter, then the above test utilizes in some sense all the relevant information about the comparison of the average growth curves.

(ii) For the application of this test, it is not necessary to know the exact values of the time points at which observations are made. It is, of course, necessary that at each time point, the observations should have been taken on all the individuals in the experiment.

(iii) There are many experiments in which the responses are obtained under different conditions or for different groups subjected to a graded set of doses not quantitatively measurable. Under the assumption that response is an increasing function of dose, the technique developed above can be used.

(iv) One can also fit a quadratic in the estimated time metameter for each individual curve and examine group differences in the second degree term and proceed to higher degree terms if necessary.

(v) Besides a transformation of the time variable it may be desirable to also transform the variable under study (for instance, to the logarithm in the case of weight) to secure a closer straight-line trend.

1.2. *Tests of further aspects of the null hypothesis concerning equality of average growth curves*

The analysis of Sec. 1.1 would result in a comparison of the growth curves in all relevant aspects only if the average curves in different groups could be made linear by a common time metameter. The hypothesis of the existence of a common transformation can, however, be subjected to a test by the procedure outlined in Sec. 2 when the sample size is large.

If there are k groups to be compared in each of p measurements representing growth in p successive given time periods, then we have $p(k-1)$ degrees of freedom out of which $(k-1)$ degrees of freedom have been used in comparing the average growth as illustrated in Sec. 1.1. To obtain a test criterion based on the rest of the degrees of freedom, we formally define a hypothesis

$$\frac{\mu_{si}}{\theta_i} - \frac{\mu_{sj}}{\theta_j} = \frac{\mu_{ri}}{\theta_i} - \frac{\mu_{rj}}{\theta_j}, \tag{2}$$

$$r, s = 1, \cdots, k; \qquad j = 1, \cdots, p,$$

where μ_{ij} is the mean of y_j for the i-th group and the θ_i are the estimates obtained in Sec. 1.1. The equations (2) can be recognised as a set of linear hypotheses which can be tested by a suitable Wilk's criterion whose computation is explained below.

We first obtain an analysis of dispersions of the variables y_1, y_2, y_3, and y_4 as between and within groups. The sum of squares and products (S.P.) matrix within groups having 24 degrees of freedom is taken and to it is appended an extra column and row containing θ_1, θ_2, θ_3, θ_4, and zero in the pivotal position as shown below:

582.3	42.5	−55.5	−74.6	θ_1
	609 0	626.5	344.5	θ_2
		1046.7	459.0	θ_3
			853.0	θ_4
				0.

In the above matrix the elements below the diagonal are omitted because of symmetry. Substituting the values $\hat{g}_1 = 603$, $\hat{g}_2 = 673$, $\hat{g}_3 = 570$, $\hat{g}_4 = 674$ obtained before, the value of the determinant Δ is computed by any standard method. Thus

$$\Delta(\text{for error}) = -1291 \times 10^{11}$$

The same procedure is repeated with the total S.P. matrix (between + within groups) having (24 + 2) degrees of freedom.

The matrix of the determinant to be evaluated is indicated by

664.0	79.7	−44.0	38.3	603
	1085.9	1409.2	1131.9	673
		2362.6	1719.1	570
			2187.0	674
				0

$$\Delta \text{ for (error + between)} = -3045 \times 10^{11}$$

The criterion is

$$\Lambda = \frac{\Delta(\text{error})}{\Delta(\text{error + between})} = \frac{1291}{3045} = 0.4241.$$

In the usual notation (see Rao [1952], p. 260), $n = 26$, $q = 2$, and $p = 3$, since there are effectively 3 variables representing successive differences of (y_i/g_i). In this case an exact test is available in the form of a variance ratio

$$F = \left[\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right]\left[\frac{n-p-1}{p}\right] = 3.93$$

with $2p = 6$ and $2(n - p - 1) = 44$ degrees of freedom. The observed value 3.93 exceeds the five percent value of F giving further evidence of differences between the growth curves under different treatments.

The device used in defining the formal hypothesis (2) is just to obtain a statistic, which records deviations from the hypothesis specifying equality of growth curves and whose distribution is exact under that null hypothesis. The equations (2) cannot define a hypothesis unless the g_i have assigned values and are not estimates. With given values of g_i, the test criterion, whose computation is explained above, is valid for testing such a hypothesis. This hypothesis would imply that, with respect to a given time metameter, the average growth curves in all groups are linear or, in other words, there is no interaction between growth rates (with respect to the chosen time scale) and successive intervals of time.

In large samples the estimates $\hat{g}_i$ tend to their expected values, in terms of which the hypothesis under consideration can be strictly defined. In such a case, when stable values of θ_i are obtained, the test criterion derived using the estimates θ_i may be interpreted as providing a test of goodness of fit of a linear trend with respect to a common time metameter or of interaction referred to above. But in finite samples, it overestimates significance as a test of this hypothesis. However, when the test does not show significance, one can conclude that there is no evidence of departure from this hypothesis.

2. ALTERNATIVE TESTS VALID FOR LARGE SAMPLES

The tests developed in Sec. 1 are simple in the sense that they do not involve complex computational techniques. More efficient tests depending on the roots of determinantal equations can be constructed but their exact distributions are not known. Fairly good approximations to the percentage points are available when the sample size is large.

Let us consider the problem of goodness of fit of linear trend with respect to a common time metameter discussed in Sec. 1.2. The following notations are used:

k = number of groups,

n_i = sample size for the i-th group,

$\bar{y}_j^{(i)}$ = the observed average growth in the j-th period for the i-th group,

$B_{rs} = \sum n_i \bar{y}_r^{(i)} \bar{y}_s^{(i)}$, the uncorrected sum of squares and products between groups,

T_{rs} = the total uncorrected (for the mean) sum of squares and products within groups; and

$S_{rs} = T_{rs} - B_{rs}$, the corrected sum of squares and products within groups.

Construct the determinantal equation

$$| S - \lambda T | = 0. \tag{3}$$

The likelihood-ratio criterion for testing the goodness of fit of linear trend for growth curves in different groups is

Λ = product of the $(p - 1)$ largest roots of equation (3)

$$= \frac{| S |}{\lambda_p | T |}$$

where λ_p is the smallest root. In large samples the statistic

$$-(\sum n_i) \log_e \Lambda \tag{4}$$

can be used as χ^2 with $(p - 1)(k - 1)$ degrees of freedom. With the multiplying coefficient suggested by Bartlett [1948], the χ^2-approximation is

$$-\left(\sum n_i - \frac{p+k}{2}\right) \log_e \Lambda \tag{5}$$

which differs from (4) in the multiplying constant.

The estimate of the common direction $(\mu_1, \cdots, \mu_p)$ which provides a transformation of the time variable when not invalidated by the above test can be obtained from the latent vector corresponding to the smallest root of equation (3). If $(h_1, h_2, \cdots, h_p)$ is the latent vector, then

$$\mu_i \propto h_1 T_{i1} + h_2 T_{i2} + \cdots + h_p T_{ip}, \quad i = 1, \cdots, p,$$

where the T_r are the elements of the matrix T. These are, perhaps, more efficient estimates than those obtained in Sec. 1.1 by averaging the observations on gain in weight for each time interval.

The likelihood-ratio criterion for testing the equality of regression coefficients (rate of growth with respect to the common time metameter) is found to be

$$\Lambda_1 = \frac{\lambda_p \mid T \mid}{\mid t \mid}, \tag{6}$$

where (t) is the matrix of corrected total sums of squares and products, while it may be recalled that (T) refers to the uncorrected sums of squares and products, and λ_p is the smallest root of equation (3) In this case we may use the variance-ratio approximation,

$$F = \left[\frac{(\sum n_i - k)}{(k - 1)}\right]\left[\frac{1 - \Lambda_1}{\Lambda_1}\right] \tag{7}$$

with $(k - 1)$ and $(\sum n_i - k)$ degrees of freedom. It may be observed that this test may be used even if the χ^2-test of (4, 5) is significant, but it acquires a special significance when a common transformation of the time axis is indicated by the Λ-test.

3. INVESTIGATION OF GROWTH MODELS

3.1. *Tests based on the dispersion matrix*

In Sec. 1 and 2, tests were developed to examine whether, by a common transformation, the average growth curves of different groups can be made linear and also to test whether the slopes are the same for all the groups. No assumption was, however, made about an individual growth curve. The observations $y_1, \cdots, y_p$, representing growth in the p time periods, were allowed to follow an arbitrary p-variate normal distribution. Modification and improvement may be desirable if something is known about the stochastic nature of growth.

We may consider the model

$$y_{t\alpha} = \lambda_\alpha g(t) + \epsilon_t \tag{8}$$

where $y_{t\alpha}$ is the increase in the t-th interval, λ_α is a parameter specific to individual α, $g(t)$ an unknown function of time only and ϵ is a random error. The errors ϵ_t and $\epsilon_{t'}$ for any two time periods are taken to be uncorrelated. It is believed that, apart from a deterministic linear trend for growth with respect to some time metameter, there are independent disturbances taking place in small intervals of time. Under such conditions the growth in any given period can be represented by the expression (8).

The model (8) implies that, by a common transformation $\tau = g(t)$, all the individual growth curves can be made linear apart from random fluctuations. It may be observed that in the analysis of previous sections this model (without the random error) was used *only* for the true average growth curves of different groups. The particular stochastic nature of individual curves described by the model (8) was not used in tests of significance.

The analogy of equation (8) with that used in factor analysis suggests the more general model

$$y_{t\alpha} = \lambda_\alpha^{(1)} g_1(t) + \lambda_\alpha^{(2)} g_2(t) + \cdots + \epsilon_t \tag{9}$$

where $\lambda^{(1)}, \lambda^{(2)}, \cdots$ correspond to different factors and $g_1, g_2, \cdots$ the regression coefficients. If such a representation is true, we should be able to replace the growth curve by its estimated factor values $\lambda^{(1)}$ $\lambda^{(2)}, \cdots$ and choose the dominant ones for further analysis.

There is, however, one difference between the model (9) and factor analytic models which include a constant $c(t)$ representing the mean of $y(t)$. For this reason, we cannot directly apply the tests used in factor analysis. One could obtain the likelihood-ratio test appropriate to test the model (9) but this appears to be complicated. As a first step the tests developed in factor analysis to determine the number of factors etc. (see Lawley [1914], Rao [1945]) can be used since the dispersion matrix of $y_1, y_2, \cdots$ is the same for both the models. The model (9) has, however, further restrictions on the mean values which have to be tested.

If in this model we replace ϵ_t by ϵ independent of t, then we can use Hotelling's principal component analysis to determine the number of factors. In this case, the test proposed by Bartlett [1950] in the investigation of factors is applicable (see Rao [1954]).

3.2. *Estimation of factors and tests of significance for differences between groups*

As in Sec. 1, let us consider the problem of comparing growth curves classified under different treatments. Let us assume that

$$y_{t\alpha} = \lambda_\alpha^{(1)} g_1(t) + \cdots + \lambda_\alpha^{(k)} g_k(t) + \epsilon$$

and estimate all λ and g by minimising the expression

$$\sum_t \sum_\alpha [y_{t\alpha} - \lambda_\alpha^{(1)} g_1(t) - \cdots - \lambda_\alpha^{(k)} g_k(t)]^2$$

which leads to the principal component analysis of the uncorrected sum of squares and products matrix. Let T represent this matrix, the typical element T of which is computed from the formula

$$T_{rs} = \sum_\alpha y_{r\alpha} y_{s\alpha}$$

(The sum over α is a sum over individuals).

Consider the determinantal equation

$$| T - \mu I | = 0$$

and find the first k latent vectors corresponding to the first k dominant roots.

Root	Latent vector
$\mu_1(\max)$	$g_1(1), g_1(2), \cdots, g_1(p)$
$\vdots$	$\vdots$
μ_k	$g_k(1), g_k(2), \cdots, g_k(p)$

The latent vectors provide the values of the functions $g_1(t), g_2(t) \cdots g_k(t)$ at the time points $1, 2, \cdots p$. For any individual the λ-values are obtained from the formulae

$$\lambda_\alpha^{(1)} = y_{1\alpha} g_1(1) + y_{2\alpha} g_1(2) + \cdots + y_{p\alpha} g_1(p)$$
$$\cdots \qquad \cdots \qquad \cdots \qquad \cdots$$
$$\lambda_\alpha^{(k)} = y_{1\alpha} g_k(1) + y_{2\alpha} g_k(2) + \cdots + y_{p\alpha} g_k(p)$$

which are linear combinations of the observations with the coefficients provided by the latent vectors. By this process the p observations on increase in weight are replaced by k values which in some sense represent the dominant aspects of the growth curve. The methods of multivariate analysis for testing the differences between treatments, etc. can now be used on the k reduced variables though the tests may not be exact. The analysis can be undertaken in the order provided by the latent vectors, first testing differences in $\lambda^{(1)}$, and then differences in $\lambda^{(2)}$ independently of $\lambda^{(1)}$, and so on, eliminating the initial weight if necessary.

4. ILLUSTRATIVE EXAMPLE OF GROWTH PROCESSES UNDER UNCONTROLLED CONDITIONS

The methods discussed in this paper were originally developed with the intention of applying them on weights of babies obtained at periodic intervals during the first year of growth. The data were collected by Dr. M. N. Rao and Dr. B. Bhattacharyya of the All India Institute of Hygiene and Public Health with the intention of setting up the norms and variations for weights of babies during the first year of growth. Their results are published in two papers (Rao and Bhattacharyya [1952] and [1953]).

The data collected by them is of great scientific interest since they provide a realistic picture of growth under natural (uncontrollable) conditions and this is what the authors were aiming at. About 100 babies, 50 boys and 50 girls, were selected and investigators visited their houses in Calcutta periodically to obtain their weights. An inevitable feature of such an investigation is incomplete records. Observations could not be continued on all children even for such a short period as one year mainly due to death, disease of the children, and, to some extent, the transfer of parents to outside stations. On searching the records it was discovered that only in 14 cases for boys and 13 for girls could the observations be continued till the end of one year. (Nearly 75 percent casualties due to death, disease, and other causes!) This is an important factor which future investigators may bear in mind.

The number of cases with complete records is not sufficiently large in this study to carry out any analysis of practical value but it would indeed be worthwhile examining how these 27 children have struggled during their first year of existence and survived to provide us with complete records. The reason for confining the analysis of the records to onlp 27 babies, omitting the incomplete records, is as follows: the characteristics of growth in the case of children with incomplete records due to death, disease, etc. will necessarily be different from those of

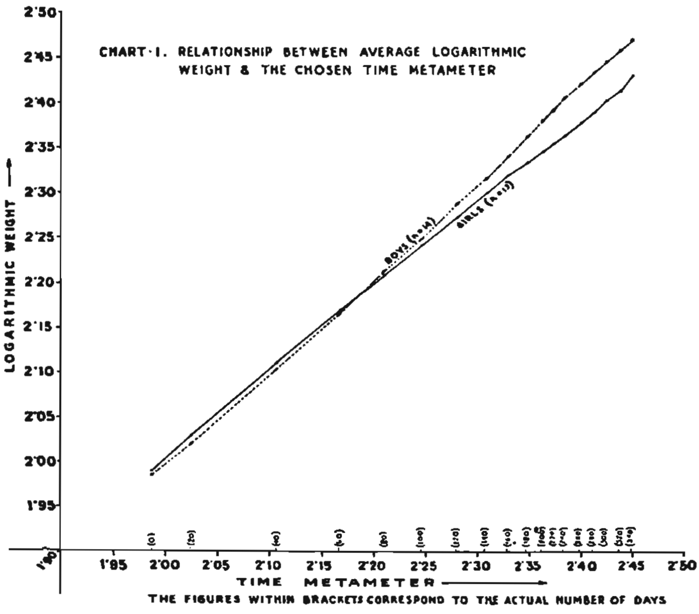
TABLE 3

MEAN VALUES OF WEIGHT (ABSOLUTE AND LOGARITHMIC) AND GAIN IN WEIGHT FOR BOYS AND GIRLS AT INTERVALS OF 20 DAYS
(Data: M. N. Rao and B. Bhattacharya)

Time in days	Absolute weight in ounces					Logarithmic weight: base 10				
	Average		Time metameter	Average gain in successive intervals		Average		Time metameter	Average gain in successive intervals	
	Boys ($n = 14$)	Girls ($n = 13$)		Boys	Girls	Boys	Girls		Boys	Girls
0	100.29	99.62	99.96	—	—	1.985	1.989	1.987	—	—
20	108.29	109.54	108.89	8.00	9.92	2.020	2.029	2.025	.024	.034
40	130.64	131.54	131.07	22.35	22.00	2.102	2.110	2.106	.082	.080
60	150.36	149.77	150.07	19.72	18.23	2.164	2.168	2.166	.062	.058
80	167.14	164.23	165.74	16.78	14.46	2.212	2.208	2.210	.048	.040
100	182.50	176.77	179.74	15.36	12.54	2.249	2.242	2.246	.037	.033
120	198.14	189.85	194.15	15.64	13.08	2.286	2.273	2.280	.037	.032
140	210.93	201.00	206.15	12.79	11.15	2.315	2.299	2.307	.029	.025
160	221.86	209.85	216.07	10.93	8.85	2.339	2.318	2.329	.023	.019
180	233.14	217.69	225.70	11.28	7.84	2.361	2.333	2.348	.023	.015
200	242.43	223.69	233.41	9.29	6.00	2.379	2.345	2.363	.018	.012
220	248.21	228.15	238.56	5.78	4.46	2.390	2.354	2.373	.011	.008
240	255.93	232.69	244.74	7.72	4.54	2.404	2.363	2.384	.014	.009
260	264.57	239.54	252.52	8.64	6.85	2.419	2.376	2.398	.015	.013
280	272.57	246.62	260.07	8.00	7.08	2.432	2.389	2.411	.013	.013
300	280.00	253.46	267.22	7.43	6.84	2.444	2.401	2.423	.012	.012
320	288.07	260.69	274.89	8.07	7.23	2.457	2.413	2.436	.012	.012
340	296.07	269.00	283.04	8.00	8.31	2.469	2.427	2.449	.012	.013

the rest, so much so that the information from the incomplete records could not be pooled with the rest either for the purpose of determining the form of the average growth curve or the stochastic nature of the individual growth curves over the entire time period.

Table 3 gives the mean values of weights (absolute and logarithmic) for boys and girls at intervals of twenty days from birth, the time metameter with respect to which the growth is expected to have a linear trend (which is estimated by the method of Rasch by averaging the weights of all the 27 children), and also values for gains in weight during the successive intervals.



An examination of the figures in Table 3 and the chart based on them leads to the following conclusions.

(i) The rate of growth in absolute weight or relative to absolute weight appears to be generally greater for boys than for girls during the first year of growth.

(ii) The rate of growth steadily decreases as time increases up to a certain time (220 days in this case uniformly for boys and girls) and then increases for a short while. This seems to demand an explanation.*

In the interpretation of the rate of growth in the first 20 days one must take into account the fact that the weight of a baby decreases within a first few days after birth and then increases.

(iii) With respect to the estimated time metameter the mean growth curves (Chart 1) for boys and girls appear to be faithfully linear. Although logarithmic weight is chosen for presentation in the chart the same is true of the absolute weights with respect to its own time metameter. Surprisingly, the individual growth curves also exhibit linear trend. They are not reproduced here for want of space. At least in the first year of growth it appears that growth is largely controlled by a single factor.

5. THE PROBLEM OF THE DISCRIMINANT FUNCTION WITH CONTINUOUS CURVES

In particular cases, if devices exist to record growth continuously with time, we have the problem of comparing the averages of entire curves and not merely at a number of time points. It may be necessary to consider the derivative curves for comparison. Our attempt here is only to develop the most general solution in the problem of discrimination when observations consist of continuous curves.

Let $f_\alpha(t)$ represent the observed curve for an individual α in the time interval, say (0, 1). The expectation curve for the group is

$$E\{f_\alpha(t)\} = f(t)$$

and the dispersion function is

$$E[f_\alpha(t)f_\alpha(t')] - f(t)f(t') = D(t, t').$$

The problem is one of determining a linear functional $L\{f(t)\}$ with respect to which two given groups differ the most. For continuous functions $f(t)$ it is known that the following integral representation of a linear functional holds

$$L\{f(t)\} = \int_0^1 f(t)\, dg(t).$$

*In India, generally, milk is the only diet given to a child for about six months. This is supplemented by rice or some other form of starch between the 7-th and 9-th months depending on the family custom.

The problem then reduces to the determination of $g(t)$.

As in the p-variate problem we will maximise the ratio of the square of the difference in mean values for two groups to the common variance. If α and β denote individuals in the two groups,

$$E[L\{f_\alpha(t)\}] - E[L\{f_\beta(t)\}] = \int_0^1 \{f_1(t) - f_2(t)\}\, dg(t)$$

$$= \int_0^1 d(t)\, dg(t),$$

where $d(t)$ is the difference between average curves. The variance of $L\{f(t)\}$ is

$$\iint D(t, t')\, dg(t)\, dg(t').$$

The ratio to be maximised is

$$\iint d(t)\, d(t')\, dg(t)\, dg(t') / \iint D(t, t')\, dg(t) dg(t').$$

The function $g(t)$ for which the above expression is a maximum is obtained as a solution of the integral equation

$$d(t) = \int D(t, t')\, dg(t').$$

There is no general method of solving this equation for any given dispersion function $D(t, t')$ except numerically by reducing it first to a problem of curves approximated by straight lines whose number is increased till $g(t)$ is determined with the requisite accuracy. This is equivalent to comparing the curves at a finite number of points each time. If special forms are assumed for the dispersion function, then direct solutions may exist. Further work done in this direction will be reported elsewhere.

REFERENCES

1. Bartlett, M. S. [1947]. Multivariate analysis. *J.R.S.S. Suppl.* 9, 76.
2. Bartlett, M. S. [1950]. Tests of significance in factor analysis. *Brit. J. Psychol. Statist. Sect.* 3, 77.
3. Box, G. E. P. [1950]. Problems in the analysis of growth and wear curves. *Biometrics* 6, 362.
4. Lawley, D. N. [1941]. Further investigations in factor estimation. *Proc. Roy. Soc. Edin.* 61, 176.
5. Rao, C. R. [1952]. *Advanced Statistical Methods in Biometric Research.* John Wiley and Sons, New York.

6. Rao, C. R. [1954]. Estimation and tests of significance in factor analysis. *Psychometrika* 20, 93.
7. Rao, M. N. and Bhattacharyya, B. [1952]. Growth curves of children in first month. *Indian Journal of Pediatrics* 19, 1.
8. Rao, M. N. and Bhattacharyya, B. [1953]. Growth curves of children in first year. *Indian Journal of Pediatrics* 20, 249.
9. Wishart, J. [1938]. Growth rate determination in nutrition studies with bacon pig and their analysis. *Biometrika* 30, 16.