IMPROVING THE ACCURACY OF GROWTH INDICES BY THE USE OF RATINGS

By G. A. McIntyre* and R. F. Williams†

[*Manuscript received May* 16, 1949]

Summary

A statistical procedure is developed whereby the precision of estimation of growth increments and various growth indices is greatly increased, especially where the variability of the plant material is great.

The procedure takes account of the fact that the difference between the mean weights of two successive harvests includes the difference between the sample means at the time of the first harvest. The importance of this factor is reduced by the use of ratings of both samples taken at the time of the first harvest. Weight comparisons are made by reference to the mean rating at this time or, where a succession of harvests is involved, to a suitable estimate of this mean rating.

The procedure is applied to a study on growth of tomatoes on a range of soil treatments and using simple chains of leaf area ratings. It is exemplified in detail from the control series of that experiment.

The data are examined critically to see whether they satisfy the assumptions inherent in the development of the theory. It is found that the variables of the bivariate distributions are highly correlated, with no evidence of a departure from a linear trend. Under these conditions, bias introduced from small departures from normality in the marginal distributions will be negligible.

Estimates of total weight, leaf weight, and leaf area based on maximum likelihood estimates of mean rating are more precise than are those based on mean rating at first harvest.

Gains in precision in estimates of relative growth rate and net assimilation rates are quite substantial, but there is little advantage in the use of maximum likelihood estimates in place of mean rating at first harvest for this purpose.

For estimates of weight, leaf area, and growth indices, the gain in information using ratings is as great for the absolute as it is for the logarithmic data.

General considerations relevant to the application of the procedure are discussed, and its merits and limitations are indicated.

I. Introduction

The conventional procedure for determining the increments in total dry weights, leaf weights, etc. of a plant species growing under a specific set of conditions is to make a succession of harvests of random samples, determine the means, and from these to estimate the growth increments and various growth indices. The difference between the mean weights of two harvests

* Section of Mathematical Statistics, C.S.I.R.O.

† Division of Plant Industry, C.S.I.R.O.; working at the C.S.I.R.O. Irrigation Research Station, Griffith, N.S.W.

involves not only the sampling variation in the increments for individual plants between the two harvests but also the deviation of the mean of the later harvested plants from the other set at the time of the first harvest. As the interval between harvests is shortened this latter factor becomes progressively more important until in the limit of zero increment it is the sole source of error.

The importance of this factor can be reduced if it is possible to make objective measurements or ratings* which do not harm the plants on both sets at the time of the first harvest, these measurements being highly correlated with total weights, leaf weight, etc. Comparisons between weights at the harvests may then be made at the mean rating, the sampling error of which is only of importance if the regression slopes of weight on rating at the two harvests diverge. The sampling error of the mean rating can be reduced if plants not harvested on either occasion are also rated at the first harvest. More than one rating could be used but the increase in computational labour would rarely justify this extension.

II. Rating for a Succession of Harvests

This principle of making comparisons over any interval by reference to a mean rating at the beginning of the interval leads directly to a variety of possible methods of arranging ratings for a succession of harvests. One of the simplest is to rate at the beginning of each interval only the sets of plants to be harvested at the beginning and end of the interval. Excluding the first and last harvests all plants for intermediate harvests are rated twice, at the beginning of the interval and at the end immediately preceding harvest.

For such intermediate harvests we may then have for each plant the initial and final ratings and the final leaf and total weights. If these measures can be regarded as distributed in a multivariate normal distribution, and this applies to the measures for each successive sample, and, further, if the variances and covariances for each of these distributions are known, one can readily estimate by the method of maximum likelihood or by least squares the population means and error of estimate for each measure at each harvest, and also any function of these parameters such as growth indices, differences in growth indices over successive intervals, etc.

In practice the population variances and covariances are not known. The errors of estimate with sample variances and covariances substituted for the corresponding population values will in consequence understate the true error by an amount which would be very laborious to estimate in any instance. For this and other reasons, including computational simplicity in developing and solving the normal equations, this approach has been confined in the subsequent treatment to improved estimates of only the mean rating at each harvest and it will be convenient to outline the procedure at this point.

* The presence of large subjective errors is not necessarily to be inferred from the use of this term. In general, the more objective the rating the better.

As an example consider the simple chain outlined above with only three harvests (Fig. 1), where $\bar{x}$ is the mean rating of a set of plants and the subscript indicates the interval and whether the measure is made at the beginning (b) or end (e). Suppose the number of cases associated with $\bar{x}_{01e}$, $\bar{x}_{12b}$, and $\bar{x}_{23b}$ are n_1, n_2, and n_3 respectively. Let the population mean ratings at harvests 1 and 2

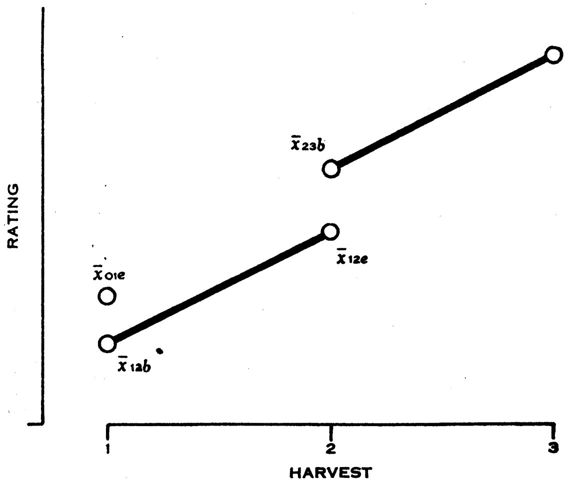

Fig. 1

be μ_{10} and μ_{01}, the variances μ_{20} and μ_{02}, the covariance μ_{11} and the correlation coefficient ρ. Then the joint sampling distribution of the sample mean ratings is

$$\text{constant} \times \exp. -\tfrac{1}{2}\Big[\frac{n_1(\bar{x}_{01e}-\mu_{10})^2}{\mu_{20}}+\frac{n_2}{1-\rho^2}\Big\{\frac{(\bar{x}_{12b}-\mu_{10})^2}{\mu_{20}} - \frac{2\mu_{11}(\bar{x}_{12b}-\mu_{10})(\bar{x}_{12e}-\mu_{01})}{\mu_{20}\,\mu_{02}}+\frac{(\bar{x}_{12e}-\mu_{01})^2}{\mu_{02}}\Big\} + \frac{n_3(\bar{x}_{23b}-\mu_{01})^2}{\mu_{02}}\Big]\,d\bar{x}_{01e}\,d\bar{x}_{12b}\,d\bar{x}_{12e}\,d\bar{x}_{23b}\,. \qquad \ldots (1)$$

Minimizing the logarithm for μ_{10} and μ_{01} gives as normal equations

$$\frac{n_1(\bar{x}_{01e}-\mu_{10})}{\mu_{20}}+\frac{n_2}{1-\rho^2}\Big\{\frac{(\bar{x}_{12b}-\mu_{10})}{\mu_{20}}-\frac{\mu_{11}(\bar{x}_{12e}-\mu_{01})}{\mu_{20}\mu_{02}}\Big\}=0,$$

$$\frac{n_2}{1-\rho^2}\Big\{-\frac{\mu_{11}(\bar{x}_{12b}-\mu_{10})}{\mu_{20}\mu_{02}}+\frac{(\bar{x}_{12e}-\mu_{01})}{\mu_{02}}\Big\}+\frac{n_3(\bar{x}_{23b}-\mu_{01})}{\mu_{02}}=0,$$

which have as solution for μ_{10},

$$\mu_{10}=\{(n_1+n_2)(n_2+n_3)\mu_{20}\mu_{02}-n_1n_3\mu_{11}{}^2\}^{-1}\,[n_2n_3(\bar{x}_{23b}-\bar{x}_{12e})\mu_{11}\mu_{20} + n_2(n_2+n_3)\bar{x}_{12b}\mu_{20}\mu_{02}+n_1\bar{x}_{01e}\{(n_2+n_3)\mu_{20}\mu_{02}-n_3\mu_{11}{}^2\}].$$

It will be noted that the sum of the coefficients of $\bar{x}_{12b}$ and $\bar{x}_{01e}$ equals unity while the coefficients of $\bar{x}_{23b}$ and $\bar{x}_{12e}$ are equal and opposite so that the expectation is unbiassed. Similar relations hold of course for μ_{01}. The usual method of numerical solution would be to form the reciprocal matrix of the coefficients of μ_{10} and μ_{01}. Then the estimate of μ_{10} is

$$c_{11}\left\{\frac{n_1\bar{x}_{01e}}{\mu_{20}}+\frac{n_2}{1-\rho^2}\left(\frac{\bar{x}_{12b}}{\mu_{20}}-\frac{\mu_{11}\bar{x}_{12e}}{\mu_{20}\mu_{02}}\right)\right\}$$
$$+c_{12}\left\{\frac{n_2}{1-\rho^2}\left(-\frac{\mu_{11}\bar{x}_{12b}}{\mu_{20}\mu_{02}}+\frac{\bar{x}_{12e}}{\mu_{02}}\right)+\frac{n_3\bar{x}_{23b}}{\mu_{02}}\right\}. \quad \ldots (2)$$

The collected coefficients of $\bar{x}_{01e}$, $\bar{x}_{12b}$, and $\bar{x}_{23b}$ should then identically satisfy the above relations. If sample estimates of μ_{20}, μ_{11}, and μ_{02} are used, and there will be different estimates of μ_{20}, for example, from x_{01e} and x_{12b}, the same relations will apply as these differences can be absorbed in the n_1,n_2,n_3 multipliers. The error of the unbiassed estimate as given by the c matrix will, however, tend to understimate the true error.

An alternative approach to the estimate of μ_{10} is to determine the weighted mean with least error of the estimates given by $\bar{x}_{01e}$, $\bar{x}_{12b}$ and $\bar{x}_{12b}+\bar{x}_{23b}-\bar{x}_{12e}$. The variances and covariances of these estimates can be estimated and hence the best weighted mean. This will give the same solution as the method outlined above. Other measures which are less efficient but unbiassed can be developed along the same lines. Thus we may take $\frac{1}{2}(\bar{x}_{01e}+\bar{x}_{12b})$ as one estimate and $\bar{x}_{12b}+\bar{x}_{23b}-\bar{x}_{12e}$ as another and determine the best weighted mean of these.

Either approach, through normal equations or weighted means, will provide estimates of the population means and their sampling errors but the latter will be biassed because the errors in the weights, that is, in the variances and covariances, are ignored.

III. Improved Estimate of Population Mean

Consider now the use of ratings to improve the estimates of the population mean weights, etc. In the subscript notation of the following sections, the number 1 will refer to plants in a first interval and 2 to plants in the following interval, both sets of plants being rated at the harvest separating the two intervals. As before, measures at the beginning and end of an interval will be denoted by (b) and (e) respectively.

Assuming that weights and ratings are bivariate normal variables then between y_{1e} and x_{1e} for example we have a regression relation of the form

$$y=\bar{y}_{1e}+b_1(x-\bar{x}_{1e}),$$

where $\bar{y}_{1e}$ and $\bar{x}_{1e}$ are the mean weights and ratings of n_1 plants. If $\hat{x}$ is the mean of these n_1 and $N-n_1$ additional ratings at the same time then

$$\hat{y}=\bar{y}_{1e}+b_1(\hat{x}-\bar{x}_{1e}) \quad \ldots (3)$$

has an expectation equal to the population mean ε and the expectation of $(\hat{y}-\varepsilon)^2$ is

$$\frac{\sigma^2_{y_{1e}}-\beta_1{}^2\sigma^2_{x_{1e}}}{n_1}+\sigma^2_{b_1}\left(\frac{\sigma^2_{x_{1e}}}{n_1}-\frac{\sigma^2_{x_{1e}}}{N}\right)+\beta_1{}^2\frac{\sigma^2_{x_{1e}}}{N}, \quad \ldots (4)$$

the components of which may be referred to as array variation, slope variation and reference point variation respectively, β_1 being the parameter of which b_1 is an estimate. Denoting

$$\frac{\Sigma(y-\bar{y}_{1e})^2-b_1{}^2\Sigma(x-x_{1e})^2}{n_1-2} \text{ as } S^2{}_{y_{1e}\cdot x_{1e}}$$

an unbiassed estimator of this variance is

$$\frac{S^2{}_{y_{1e}\cdot x_{1e}}}{n_1}+\frac{S^2{}_{y_{1e}\cdot x_{1e}}}{n_1-3}\left(\frac{1}{n_1}-\frac{1}{N}\right)+\frac{1}{N}\left(\frac{b_1{}^2\Sigma(x-\bar{x}_{1e})^2}{n_1-2}-\frac{\Sigma(y-\bar{y}_{1e})^2}{(n_1-1)(n_1-2)}\right),$$

which can be alternatively written

$$S^2{}_{y_{1e}\cdot x_{1e}}\left\{\frac{n_1-2}{n_1(n_1-3)}-\frac{2n_1-4}{N(n_1-1)(n_1-3)}\right\}+\frac{b_1{}^2\Sigma(x-\bar{x}_{1e})^2}{N(n_1-1)} \quad . . (5)$$

If for $\hat{x}$ we had substituted the maximum likelihood estimate for the rating at this harvest with variance $S^2{}_{\hat{x}}$, an approximate estimate of the variance of the estimate $\hat{y}_{1e}$ is given by

$$S^2{}_{y_{1e}\cdot x_{1e}}\left\{\frac{n_1-2}{n_1(n_1-3)}-\frac{2S^2\hat{x}}{\Sigma(x-\bar{x}_{1e})^2}\right\}+b_1{}^2S^2{}_{\hat{x}}$$

$$+\,2b_1\left\{\frac{\Sigma(y-\bar{y}_{1e})(x-\bar{x}_{1b})-b_1\Sigma(x-\bar{x}_{1e})(x-\bar{x}_{1b})}{n_1(n_1-2)}\times \text{coefficient of } \bar{x}_{1b} \text{ in the estimate of } \hat{x}\right\}. \quad . . . . (6)$$

The last term arises from the contribution of $\bar{x}_{1b}$ to the estimate $\hat{x}$ and the correlation of $\bar{x}_{1e}$ and $\bar{y}_{1e}$ with it. A convenient condensed notation for the factor multiplying $2b_1$ is $[y_{1e}]$. The corresponding factor of $2b_2$ in the estimate of the variance of $\hat{y}_{2e}$ referred to x_{2b} is

$$\frac{\Sigma(y-\bar{y}_{2e})(x-\bar{x}_{2e})-b_2\Sigma(x-\bar{x}_{2b})(x-\bar{x}_{2e})}{n_2(n_2-2)}\times \text{coefficient of } \bar{x}_{2e} \text{ in the estimate}$$

of $\hat{x}$ and would be designated $[y_{2e}]$.

IV. Improved Estimate of Mean Increment

Suppose now we have two normal bivariate populations which have a common marginal distribution of one of the variables x and samples of size n_1 and n_2 are drawn from them. In general the slopes of the regression lines and the dispersions about the lines for the two distributions will be different. The estimated difference of the mean y variates for the two populations at a reference point $\hat{x}$ is

$$\hat{y}_{2e}-\hat{y}_{1e}=\bar{y}_{2e}-\bar{y}_{1e}+b_2(\hat{x}-\bar{x}_{2b})-b_1(\hat{x}-\bar{x}_{1e}) . \quad . . (7)$$

If $\hat{x}$ is the mean of the n_1 and n_2 observations and a further n_3 observations on the x variate alone, when $n_1+n_2+n_3=N$, then for repeated samplings of the three sets of observations the expected value of $\hat{y}_{2e}-\hat{y}_{1e}$ is the difference in the population means for the variates, $\varepsilon_2-\varepsilon_1$, and the expectation of

$$\{(\hat{y}_{2e}-\hat{y}_{1e})-(\varepsilon_2-\varepsilon_1)\}^2 \text{ is}$$

$$\frac{\sigma^2{}_{y_{2e}\cdot x_{2b}}}{n_2}+\frac{\sigma^2{}_{y_{1e}\cdot x_{1e}}}{n_1}+\sigma^2{}_{b_2}\left(\frac{\sigma^2{}_x}{n_2}-\frac{\sigma^2{}_x}{N}\right)+\sigma^2{}_{b_1}\left(\frac{\sigma^2{}_x}{n_1}-\frac{\sigma^2{}_x}{N}\right)$$
$$+(\beta_2-\beta_1)^2\frac{\sigma^2{}_x}{N}, \quad \ldots\ldots\ldots\ldots \quad (8)$$

where $\sigma^2{}_{x_{1e}}=\sigma^2{}_{x_{2b}}=\sigma^2{}_x$.

An unbiassed estimator of this is

$$S^2{}_{y_{2e}\cdot x_{2b}}\left\{\frac{n_2-2}{n_2(n_2-3)}-\frac{2n_2-4}{N(n_2-1)(n_2-3)}\right\}+S^2{}_{y_{1e}\cdot x_{1e}}\left\{\frac{n_1-2}{n_1(n_1-3)}\right.$$
$$\left.-\frac{2n_1-4}{N(n_1-1)(n_1-3)}\right\}+\frac{1}{N}\left\{b_2{}^2\frac{\Sigma(x-\bar{x}_{2b})^2}{n_2-1}\right.$$
$$\left.-2b_2b_1\left(\frac{\Sigma(x-\bar{x}_{2b})^2\,\Sigma(x-\bar{x}_{1e})^2}{(n_2-3/2)(n_1-3/2)}\right)^{\frac{1}{2}}+b_1{}^2\frac{\Sigma(x-\bar{x}_{1e})^2}{n_1-1}\right\}. \quad \ldots \quad (9)$$

If for $\hat{x}$ the maximum likelihood estimate is used, an estimate of the variance of $\hat{y}_2-\hat{y}_1$ is given by

$$S^2{}_{y_{2e}\cdot x_{2b}}\left\{\frac{n_2-2}{n_2(n_2-3)}-\frac{2S^2{}_{\hat{x}}}{\Sigma(x-\bar{x}_{2b})^2}\right\}+S^2{}_{y_{1e}\cdot x_{1e}}\left\{\frac{n_1-2}{n_1(n_1-3)}\right.$$
$$\left.-\frac{2S^2{}_{\hat{x}}}{\Sigma(x-\bar{x}_{1e})^2}\right\}+(b_2-b_1)^2S^2{}_{\hat{x}}+2(b_2-b_1)\{[y_{2e}]-[y_{1e}]\}. \quad \ldots \quad (10)$$

The effect on the form of the expression for the variance as a result of extending from one dependent variable to two is obvious and further extension to more complex expressions than $\hat{y}_2-\hat{y}_1$ involves no particular difficulties. The only additional issue raised is in the case of an expression of the form $f(y_1, y_2)$ where y_1 and y_2 are two measures of the same plant which are correlated and which are referred to the same set of ratings. In this instance the elements of array variation are not independent and their covariance can be expressed in the form $\sigma_{y_2\cdot x}\sigma_{y_1\cdot x}\rho_{y_2y_1\cdot x}$ where $\rho_{y_2y_1\cdot x}$ is the partial correlation of y_1 and y_2 holding x constant. Similarly the covariance of the elements of slope variation is $\sigma_{b_{y_2x}}\sigma_{b_{y_1x}}\rho_{y_2y_1\cdot x}$.

V. Application to Derived Growth Indices

Consider now the application of the foregoing to growth data and derived measures such as relative growth rate and net assimilation rate. Weights of whole plants or tops of whole plants will be indicated by W, weights of leaves by LW, and leaf areas by LA. Used as ratings the leaf areas will be designated M with the mean of N simultaneous ratings or the maximum likelihood estimate taken over successive links in the chain by $\hat{M}$. The subscript notation is the same as for Section IV above.

Typical expressions for the estimate of weight are

$$\hat{W}_{1e} = \overline{W}_{1e} + b_{W_{1e}M_{1e}} (\hat{M} - M_{1e}), \quad \ldots\ldots (11)$$

and for the variance of estimate by substitution in (4), thus

$$\sigma^2_{\hat{W}_{1e}} = \sigma^2_{\overline{W}_{1e}\cdot M_{1e}} + \sigma^2_{b_{W_{1e}M_{1e}}} \left(\sigma^2_{\overline{M}_{1e}} - \sigma^2_{\hat{M}}\right) + \beta^2_{W_{1e}M_{1e}} \sigma^2_{\hat{M}} \quad \ldots (12)$$

Using for mean rating the maximum likelihood estimate, the corresponding variance from (6) is

$$S^2_{W_{1e}\cdot M_{1e}} \left(\frac{n_1 - 2}{n_1(n_1 - 3)} - \frac{2S^2_{\hat{M}}}{\Sigma(M - \overline{M}_{1e})^2}\right) + b^2_{W_{1e}M_{1e}} S^2_{\hat{M}}$$
$$+ 2b_{W_{1e}M_{1e}} [W_{1e}]. \quad \ldots\ldots\ldots\ldots (13)$$

If logarithms (to base 10) of all measures have been used instead of the actual values the expression for $S^2_{\hat{W}_{1e}}$ in logarithmic units will be as above. If now the antilog of $\hat{W}_{1e}$ be determined the corresponding variance to a good approximation will be given by

$$(\text{Antilog } \hat{W}_{1e})^2 \{S^2_{\hat{W}_{1e}} (log_e 10)^2 + \tfrac{3}{4} S^4_{\hat{W}_{1e}} (log_e 10)^4\}, \quad \ldots\ldots (14)$$

and in general the term involving $S^4_{\hat{W}_{1e}}$ can be ignored.

(*a*) *The Variance of the Relative Growth Rate*, $R = \dfrac{log_e(\overline{W}_{2e}/\overline{W}_{1e})}{t_2 - t_1}$

(i) *Without Use of Ratings.*—Arithmetic mean values are entered into the expression for R. The ratio of mean values will not be the same as the mean of the ratios for individual plants which of course cannot be determined, but will not differ substantially when changes over the interval are approximately constant multiples of the corresponding initial values for all plants in the treatment.

In the development of this and subsequent expressions it will be assumed that coefficients of variation are sufficiently small to justify the approximate methods employed.

The change in R corresponding to small changes $d\overline{W}_{2e}$ and $d\overline{W}_{1e}$ is $\dfrac{1}{t_2 - t_1}\left\{\dfrac{d\overline{W}_{2e}}{\overline{W}_{2e}} - \dfrac{d\overline{W}_{1e}}{\overline{W}_{1e}}\right\}$ so that approximately the variance of R is estimated by

$$\frac{1}{(t_2 - t_1)^2}\left\{\frac{\sigma^2_{\overline{W}_{2e}}}{\overline{W}_{2e}^2} + \frac{\sigma^2_{\overline{W}_{1e}}}{\overline{W}_{1e}^2}\right\} \ldots\ldots\ldots\ldots (15)$$

(ii) *With Use of Ratings.*—$R = \dfrac{log_e(\hat{W}_{2e}/\hat{W}_{1e})}{t_2 - t_1}$. With ratings the variation in W_2, for example, when referred to a mean of N simultaneous ratings may be regarded as

$$\Delta(\text{array } \overline{W}_{2e}) + (\overline{M}_{2b} - \hat{M})\Delta b_{W_{2e}M_{2b}} + \Delta\hat{M}\beta_{W_{2e}M_{2b}},$$

and the variation in R as

$$\frac{1}{t_2 - t_1}[\{\Delta(\text{array } \overline{W}_{2e}) + (\overline{M}_{2b} - \hat{M})\Delta b_{W_{2e}M_{2b}} + \Delta\hat{M}\beta_{W_{2e}M_{2b}}\}1/\hat{W}_{2e}$$
$$- \{\Delta(\text{array } \overline{W}_{1e}) + (\overline{M}_{1e} - \hat{M})\Delta b_{W_{1e}M_{1e}} + \Delta\hat{M}\beta_{W_{1e}M_{1e}}\}1/\hat{W}_{1e}].$$

The required variance is then

$$\frac{1}{(t_2 - t_1)^2}\left[\left\{\sigma^2_{\overline{W}_{2e}\cdot M_{2b}} + (\sigma^2_{\overline{M}_{2b}} - \sigma^2_{\hat{M}})\sigma^2_{b_{W_{2e}M_{2b}}}\right\}\frac{1}{\hat{W}_{2e}^2}\right.$$
$$+ \left\{\sigma^2_{\overline{W}_{1e}\cdot M_{1e}} + (\sigma^2_{\overline{M}_{1e}} - \sigma^2_{\hat{M}})\sigma^2_{b_{W_{1e}M_{1e}}}\right\}\frac{1}{\hat{W}_{1e}^2}$$
$$\left. + \sigma^2_{\hat{M}}\left(\frac{\beta_{W_{2e}M_{2b}}}{\hat{W}_{2e}} - \frac{\beta_{W_{1e}M_{1e}}}{\hat{W}_{1e}}\right)^2\right]. \quad \ldots\ldots\ldots (16)$$

The estimator in the case where there are N simultaneous observations at the reference harvest is obvious from preceding samples. Using the maximum likelihood reference rating the estimator is approximately

$$\frac{1}{(t_2 - t_1)^2}\left\{\frac{S^2_{W_{2e}\cdot M_{2b}}}{\hat{W}_{2e}^2}\left(\frac{n_2 - 2}{n_2(n_2 - 3)} - \frac{2S^2_{\hat{M}}}{\Sigma(M - \overline{M}_{2b})^2}\right)\right.$$
$$+ \frac{S^2_{W_{1e}\cdot M_{1e}}}{\hat{W}_{1e}^2}\left(\frac{n_1 - 2}{n_1(n_1 - 3)} - \frac{2S^2_{\hat{M}}}{\Sigma(M - \overline{M}_{1e})^2}\right) + S^2_{\hat{M}}\left(\frac{b_{W_{2e}M_{2b}}}{\hat{W}_{2e}} - \frac{b_{W_{1e}M_{1e}}}{\hat{W}_{1e}}\right)^2$$
$$\left. + 2\left(\frac{b_{W_{2e}M_{2b}}}{\hat{W}_{2e}} - \frac{b_{W_{1e}M_{1e}}}{\hat{W}_{1e}}\right)\left(\frac{[W_{2e}]}{\hat{W}_{2e}} - \frac{[W_{1e}]}{\hat{W}_{1e}}\right)\right\}. \quad \ldots (17)$$

(iii) *Logarithms Without Use of Ratings.*—If logs of leaf area ratings and of total weights have been used throughout, write $\log_e W_{2e} = U_{2e}$, $\log W_{1e} = U_{1e}$. Then the variance of R is

$$\frac{1}{(t_2 - t_1)^2}\left(\sigma^2_{\overline{U}_{2e}} + \sigma^2_{\overline{U}_{1e}}\right). \quad \ldots\ldots\ldots\ldots (18)$$

If logs to base 10 have been used this is to be multiplied by $(\log_e 10)^2$.

(iv) *Logarithms With Use of Ratings.**—The variance of R is approximately

$$\frac{1}{(t_2 - t_1)^2}\left\{\sigma^2_{\overline{U}_{2e}\cdot M_{2b}} + (\sigma^2_{\overline{M}_{2b}} - \sigma^2_{\hat{M}})\sigma^2_{b_{U_{2e}M_{2b}}} + \sigma^2_{\overline{U}_{1e}\cdot M_{1e}}\right.$$
$$\left. + (\sigma^2_{\overline{M}_{1e}} - \sigma^2_{\hat{M}})\sigma^2_{b_{U_{1e}M_{1e}}} + \sigma^2_{\hat{M}}(\beta_{U_{2e}M_{2b}} - \beta_{U_{1e}M_{1e}})^2\right\}. \quad \ldots (19)$$

* The same symbols are used for log ratings as for ratings.

(*b*) *The Variance of the Net Assimilation Rate,*

$$E_{LW} = \frac{\overline{W}_{2e} - \overline{W}_{1e}}{\overline{LW}_{2e} - \overline{LW}_{1e}} \quad \frac{\log_e(\overline{LW}_{2e}/\overline{LW}_{1e})}{t_2 - t_1}$$

(i) *Without Use of Ratings.*—The change in E_{LW} corresponding to small changes in estimated mean weights and leaf weights, $d\overline{W}_{2e}$, $d\overline{W}_{1e}$, $d\overline{LW}_{2e}$, and $d\overline{LW}_{1e}$ is

$$\frac{1}{(t_2 - t_1)(\overline{LW}_{2e} - \overline{LW}_{1e})}\left\{(d\overline{W}_{2e} - d\overline{W}_{1e})A + d\overline{LW}_{2e}\,B - d\overline{LW}_{1e}\,C\right\}$$

where $A = \log_e(\overline{LW}_{2e}/\overline{LW}_{1e})$,

$$B = \frac{\overline{W}_{2e} - \overline{W}_{1e}}{\overline{LW}_{2e}} - \frac{\overline{W}_{2e} - \overline{W}_{1e}}{\overline{LW}_{2e} - \overline{LW}_{1e}} \log_e(\overline{LW}_{2e}/\overline{LW}_{1e}),$$

$$C = \frac{\overline{W}_{2e} - \overline{W}_{1e}}{\overline{LW}_{1e}} - \frac{\overline{W}_{2e} - \overline{W}_{1e}}{\overline{LW}_{2e} - \overline{LW}_{1e}} \log_e(\overline{LW}_{2e}/\overline{LW}_{1e}).$$

Then the variance of E_{LW} is

$$\frac{1}{(t_2 - t_1)^2(LW_{2e} - LW_{1e})^2}\left\{A^2\sigma^2_{\overline{W}_{2e}} + 2AB\sigma_{\overline{W}_{2e}}\sigma_{\overline{LW}_{2e}}\rho_{W_{2e}LW_{2e}} + B^2\sigma^2_{\overline{LW}_{2e}} + A^2\sigma^2_{\overline{W}_{1e}} + 2AC\sigma_{\overline{W}_{1e}}\sigma_{\overline{LW}_{1e}}\rho_{W_{1e}LW_{1e}} + C^2\sigma^2_{\overline{LW}_{1e}}\right\}. \quad (20)$$

(ii) *With Use of Ratings.*—W_{2e} and LW_{2e} are referred to a common rating scale and similarly for W_{1e} and LW_{1e}. The change in E corresponding to changes possible in $\hat{W}_{2e}$, $\hat{W}_{1e}$, $\widehat{LW}_{2e}$, and $\widehat{LW}_{1e}$ can be symbolized by

$$\Big[\{\Delta(\text{array } \overline{W}_{2e}) + (\overline{M}_{2b} - \hat{M})\Delta b_{W_{2e}M_{2b}} + \Delta\hat{M}\beta_{W_{2e}M_{2b}}\}A$$
$$- \{\Delta(\text{array } \overline{W}_{1e}) + (\overline{M}_{1e} - \hat{M})\Delta b_{W_{1e}M_{1e}} + \Delta\hat{M}\beta_{W_{1e}M_{1e}}\}A$$
$$+ \{\Delta(\text{array } \overline{LW}_{2e}) + (\overline{M}_{2b} - \hat{M})\Delta b_{LW_{2e}M_{2b}} + \Delta\hat{M}\beta_{LW_{2e}M_{2b}}\}B$$
$$- \{\Delta(\text{array } \overline{LW}_{1e}) + (\overline{M}_{1e} - \hat{M})\Delta b_{LW_{1e}M_{1e}} + \Delta\hat{M}\beta_{LW_{1e}M_{1e}}\}C\Big],$$

divided by $(t_2 - t_1)(\widehat{LW}_{2e} - \widehat{LW}_{1e})$,

where A, B, C are as given previously but with the regression estimates of $\hat{W}_{2e}$ etc. substituted for $\overline{W}_{2e}$ etc.

The approximate variance of E_{LW} is then

$$\Big[A^2\sigma^2_{\overline{W}_{2e}\cdot M_{2b}} + 2AB\sigma_{\overline{W}_{2e}\cdot M_{2b}}\sigma_{\overline{LW}_{2e}\cdot M_{2b}}\rho_{W_{2e}LW_{2e}\cdot M_{2b}} + B^2\sigma^2_{\overline{LW}_{2e}\cdot M_{2b}}$$
$$+ \left\{A^2\sigma^2_{b_{W_{2e}M_{2b}}} + 2AB\sigma_{b_{W_{2e}M_{2b}}}\sigma_{b_{LW_{2e}M_{2b}}}\rho_{W_{2e}LW_{2e}\cdot M_{2b}} + B^2\sigma^2_{b_{LW_{2e}}}\right\} \times (\sigma^2_{\overline{M}_{2b}} - \sigma^2_{\hat{M}})$$

$$+ A^2\sigma^2_{\overline{W}_{1e}\cdot M_{1e}} + 2AC\sigma_{\overline{W}_{1e}\cdot M_{1e}}\sigma_{\overline{LW}_{1e}\cdot M_{1e}}\rho_{W_{1e}LW_{1e}\cdot M_{1e}} + C^2\sigma^2_{\overline{LW}_{1e}\cdot M_{1e}}$$

$$+ \left\{A^2\sigma^2_{b_{W_{1e}M_{1e}}} + 2AC\sigma_{b_{W_{1e}M_{1e}}}\sigma_{b_{LW_{1e}M_{1e}}}\rho_{W_{1e}LW_{1e}\cdot M_{1e}} + C^2\sigma^2_{b_{LW_{1e}M_{1e}}}\right\}$$

$$\times\, (\sigma^2_{\overline{M}_{1e}} - \sigma^2_{\hat{M}})$$

$$+ \sigma^2_{\hat{M}}\left\{A(\beta_{W_{2e}M_{2e}} - \beta_{W_{1e}M_{1e}}) + B\beta_{LW_{2e}M_{2e}} - C\beta_{LW_{1e}M_{1e}}\right\}^2 \Big]$$

$$\text{divided by } (t_2 - t_1)^2(\hat{LW}_{2e} - \hat{LW}_{1e})^2. \quad \ldots\ldots \ (\mathbf{21})$$

In practice, $\sigma_{\overline{W}_{2e}\cdot M_{2b}}\sigma_{\overline{LW}_{2e}\cdot M_{2b}}\rho_{W_{2e}LW_{2e}\cdot M_{2b}}$ would be estimated as

$$\frac{1}{n_2(n_2-2)}\left\{\Sigma W_{2e}LW_{2e} - \frac{\Sigma W_{2e}M_{2b}\Sigma LW_{2e}M_{2b}}{\Sigma M^2_{2b}}\right\} \text{ i.e. } \frac{S_{W_{2e}LW_{2e}\cdot M_{2b}}}{n_2},$$

where squares and products refer to deviates from sample means. For the maximum likelihood value of $\hat{M}$, the estimator would be

$$\Big[\Big(A^2S^2_{W_{2e}\cdot M_{2b}} + 2ABS_{W_{2e}LW_{2e}\cdot M_{2b}} + B^2S^2_{LW_{2e}\cdot M_{2b}}\Big)$$

$$\times\left(\frac{n_2-2}{n_2(n_2-3)} - \frac{2S^2_{\hat{M}}}{\Sigma(M-\overline{M}_{2b})^2}\right)$$

$$+ \Big(A^2S^2_{W_{1e}\cdot M_{1e}} + 2ACS_{W_{1e}LW_{1e}\cdot M_{1e}} + C^2S^2_{LW_{1e}\cdot M_{1e}}\Big)$$

$$\times\left(\frac{n_1-2}{n_1(n_1-3)} - \frac{2S^2_{\hat{M}}}{\Sigma(M-\overline{M}_{1e})^2}\right)$$

$$+ S^2_{\hat{M}}\left\{A(b_{W_{2e}M_{2b}} - b_{W_{1e}M_{1e}}) + Bb_{LW_{2e}M_{2b}} - Cb_{LW_{1e}M_{1e}}\right\}^2$$

$$+ 2\left\{A(b_{W_{2e}M_{2b}} - b_{W_{1e}M_{1e}}) + Bb_{LW_{2e}M_{2b}} - Cb_{LW_{1e}M_{1e}}\right\}$$

$$\left\{A([W_{2e}] - [W_{1e}]) + B[LW_{2e}] - C[LW_{1e}]\right\}\Big]$$

$$\text{divided by } (t_2 - t_1)^2(\hat{LW}_{2e} - \hat{LW}_{1e})^2. \quad \ldots\ldots \ (\mathbf{22})$$

(iii) *Logarithms Without Use of 'Ratings.*—If logarithms have been used throughout for total weights, leaf weights, and leaf areas, and we write

$$\log_e W_{2e} = U_{2e}, \qquad \log_e W_{1e} = U_{1e},$$
$$\log_e LW_{2e} = V_{2e}, \qquad \log LW_{1e} = V_{1e};$$

$$\text{then } E_{LW} = \frac{e^{\overline{U}_{2e}} - e^{\overline{U}_{1e}}}{e^{\overline{V}_{2e}} - e^{\overline{V}_{1e}}}\ \frac{\overline{V}_{2e} - \overline{V}_{1e}}{t_2 - t_1},$$

geometric means being substituted for arithmetic means in Section (i).

The change in E_{LW} corresponding to small changes in $d\overline{U}_{2e}$, $d\overline{U}_{1e}$, $d\overline{V}_{2e}$ and $d\overline{V}_{1e}$ is

$$\frac{1}{(t_2-t_1)(e^{\overline{V}_{2e}}-e^{\overline{V}_{1e}})}\Bigg[d\overline{U}_{2e}e^{\overline{U}_{2e}}(\overline{V}_{2e}-\overline{V}_{1e})-d\overline{U}_{1e}e^{\overline{U}_{1e}}(\overline{V}_{2e}-\overline{V}_{1e})$$

$$+\,d\overline{V}_{2e}e^{\overline{V}_{2e}}\left\{\frac{e^{\overline{U}_{2e}}-e^{\overline{U}_{1e}}}{e^{\overline{V}_{2e}}}-\frac{e^{\overline{U}_{2e}}-e^{\overline{U}_{1e}}}{e^{\overline{V}_{2e}}-e^{\overline{V}_{1e}}}(\overline{V}_{2e}-\overline{V}_{1e})\right\}$$

$$+\,d\overline{V}_{1e}e^{\overline{V}_{1e}}\left\{\frac{e^{\overline{U}_{2e}}-e^{\overline{U}_{1e}}}{e^{\overline{V}_{1e}}}-\frac{e^{\overline{U}_{2e}}-e^{\overline{U}_{1e}}}{e^{\overline{V}_{2e}}-e^{\overline{V}_{1e}}}(\overline{V}_{2e}-\overline{V}_{1e})\right\}\Bigg],$$

which may be written

$$\left[d\overline{U}_{2e}(\overline{W}_{2e}A)-d\overline{U}_{1e}(\overline{W}_{1e}A)+d\overline{V}_{2e}(\overline{LW}_{2e}B)-d\overline{V}_{1e}(\overline{LW}_{1e}C)\right]$$

divided by $(t_2-t_1)(\overline{LW}_{2e}-\overline{LW}_{1e})$.

The expression in Section (i) will then apply if $(\overline{W}_{2e}A)$ is substituted for A, to be used in conjunction with the variation in $\overline{U}_{2e}$ and so on. The A, B, C values of this section are of course based on geometric means.

(iv) *Logarithms With Use of Ratings.*—The development follows along the same lines as in Section (ii) with the substitutions for A etc. as given in Section (iii).

(*c*) *The Variance of the Net Assimilation Rate,*

$$E_{LA}=\frac{\overline{W}_{2e}-\overline{W}_{1e}}{\overline{LA}_{2e}-\overline{LA}_{1e}}\;\frac{\log_e(\overline{LA}_{2e}/\overline{LA}_{1e})}{t_2-t_1}$$

(i) *Without Use of Ratings.*—The required variance can be obtained from (**20**), substituting LA for LW.

(ii) *With Use of Ratings.*—In this case leaf area is being used in a dual role.

$\beta_{LA_{1e}M_{1e}}$ and $b_{LA_{1e}M_{1e}}$ each equals one, while $\sigma^2 b_{LA_{1e}M_{1e}}$, $\sigma^2\overline{LA}_{1e}.M_{1e}$, and $\rho_{W_{1e}LA_{1e}.M_{1e}}$ are zero. $[LA_{1e}]$ is also zero.

The required variance and maximum likelihood estimator can be obtained by making the obvious substitutions in (**21**) and (**22**).

(iii and iv) *Logarithms Without and With Use of Ratings.*—These follow along the same lines as (iii), (iv) of the previous section, with obvious substitutions.

VI. Application of Theory to a Specific Set of Data

(a) *Experimental and General*

In a study on growth of tomatoes on a range of soil treatments, the procedure outlined here was followed for a simple chain of leaf-area ratings. A series of photographic standards was used to estimate the areas of individual leaves and hence the leaf area for the whole plant. For the first and second harvests, very large numbers of plants were available and these were sampled thoroughly for each of the treatments. For each sample the leaf blades, including cotyledonary leaves, were separated from the rest of the shoots, and dry weights were obtained for each fraction separately.

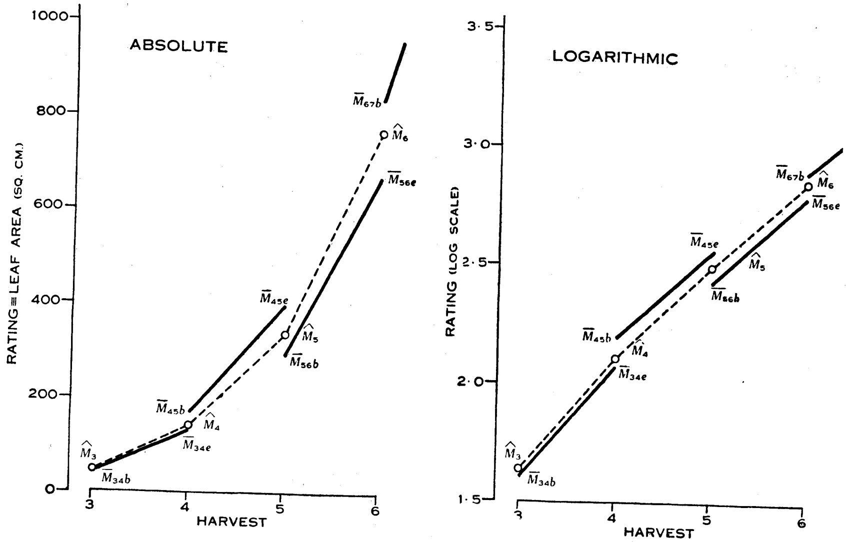

Fig. 2.—Absolute and logarithmic ratings (leaf areas) plotted against time. For plants of harvest 4, $\overline{M}_{34b}$ is the mean rating at the beginning of harvest interval 3-4, and $\overline{M}_{34e}$ is the mean rating taken at the end of the interval and just prior to harvesting the sample; $\hat{M}_4$ is the maximum likelihood estimate for the rating at harvest 4, and so on. The mean rating $\overline{M}_{23e}$ is omitted from this figure.

For harvest 3 and within each treatment, three sets of sixteen plants were taken at random and rated for leaf area. One of these sets was harvested immediately* and the dry weights of the leaf blades and total shoots determined as before.

For harvest 4, the second of these sets was again rated for leaf area and then* harvested. The third set of sixteen plants had been rated as an insurance against casualties among plants of the second group. At the time of re-rating the second set of plants, a further set of sixteen plants (including all remaining spares of the harvest 3 rating) was rated together with eight instead of sixteen plants.

* Actually, the following day in each case.

The general procedure for harvest 4 was repeated for harvests 5 and 6. At harvest 7 no estimates of leaf area were made. The first five harvest intervals were 7 days and the last interval was 14 days.

The successive ratings of the above procedure formed a chain as illustrated in Figure 2 from the absolute and logarithmic data of the control series. The picture is a simple extension of the specific case of Section II above, though the notation is slightly different. The mean rating of a set of plants is now $\overline{M}$; as before, the subscript indicates the harvest interval and whether the rating is made at the beginning (b) or end (e) of the interval. The broken lines of Figure 2 join the maximum likelihood estimates, $\hat{M}_3$-$\hat{M}_6$ of the population mean ratings for successive harvests. These give inter-harvest trends which are closely parallel to the experimental trends established by the individual links of the chain.

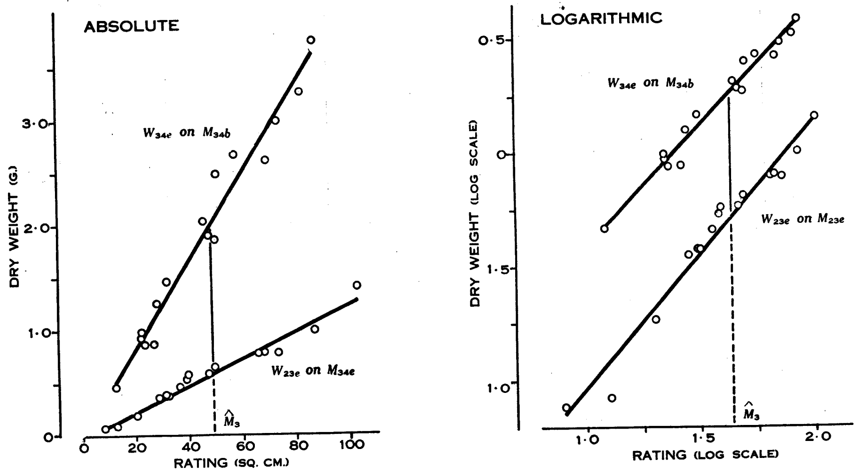

Fig. 3.—The relations between dry weight of plants (tops only) at harvests 3 and 4 and their ratings (leaf areas) at the commencement of harvest interval 3-4. The pairs of regression lines are linked by the maximum likelihood estimates, $\hat{M}_3$ of the true rating as ordinate.

The weight-rating relation (bivariate distribution) is illustrated in Figures 3-6 for the successive harvest intervals of the control series. In each case the absolute and logarithmic data are shown side by side for comparison. In preparing the figures, the log scales were kept constant throughout, but the absolute scales were scaled up or down so as to give regressions which were readily comparable with the corresponding logarithmic regressions. The weight-area scale-ratio was kept constant for all four sets of absolute data. In Figure 3, the dry weights of the plants (tops only) for harvests 3 and 4 are plotted against their leaf areas at the commencement of the interval. The two groups of sixteen weight values are highly correlated with their leaf areas, and the relations are adequately described by the linear regressions.

$$W_{23e} = -.02379 + .012575\, M_{23e},$$
$$W_{34e} = .00724 + .041194\, M_{34b},$$

and

$$\log_{10} W_{23e} = -2.2091 + 1.1687 \log_{10} M_{23e},$$
$$\log_{10} W_{34e} = -1.4490 + 1.0378 \log_{10} M_{34b}.$$

It is obvious that the departures of individual values from the regressions are trifling by comparison with their departures from the means of their marginal distributions, and it is mainly upon this fact that the whole procedure depends for the improvement of the accuracy of plant weights and growth indices.

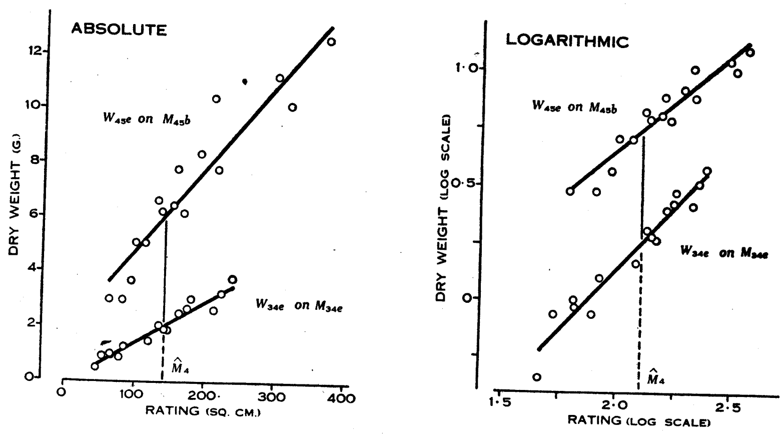

Fig. 4.—The relations between dry weight of plants (tops only) at harvests 4 and 5 and their ratings (leaf areas) at the commencement of harvest interval 4-5. The pairs of regression lines are linked by the maximum likelihood estimates, $\hat{M}_4$ of the true rating as ordinate.

The linear regressions of Figures 4-6 are as follows:

Figure 4

$$W_{34e} = -.13048 + .015202\, M_{34e},$$
$$W_{45e} = 1.74395 + .030703\, M_{45b},$$

and

$$\log_{10} W_{34e} = -2.0411 + 1.0880 \log_{10} M_{34e},$$
$$\log_{10} W_{45e} = -1.0245 + .8394 \log_{10} M_{45b}.$$

Figure 5

$$W_{45e} = -.08766 + .018185\, M_{45e},$$
$$W_{56e} = .29022 + .038189\, M_{56b},$$

and

$$\log_{10} W_{45e} = -1.8427 + 1.0364 \log_{10} M_{45e},$$
$$\log_{10} W_{56e} = -1.3644 + .9828 \log_{10} M_{56b}.$$

Figure 6

$$W_{56e} = -.42254 + .017777\, M_{56e},$$
$$W_{67e} = 49.3623 + .027836\, M_{67b},$$

and

$$\log_{10} W_{56e} = -1.9931 + 1.0792 \log_{10} M_{56e},$$
$$\log_{10} W_{67e} = .9952 + .2979 \log_{10} M_{67b}.$$

Estimates of population mean weights and hence of the mean increments from harvest to harvest are given by substituting the values of the estimates

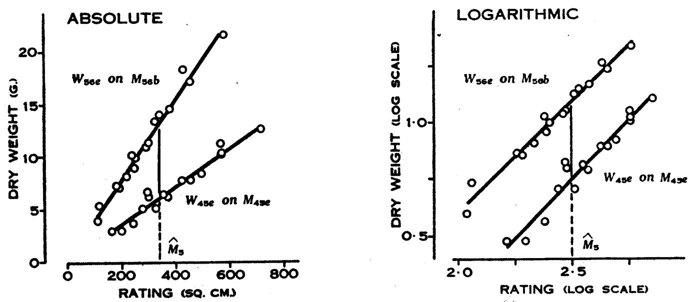

Fig. 5.—The relations between dry weight of plants (tops only) at harvests 5 and 6 and their ratings (leaf areas) at the commencement of harvest interval 5-6. The pairs of regression lines are linked by the maximum likelihood estimates, $\hat{M}_5$ of the true rating as ordinate.

of the mean ratings (see Fig. 2 and Table 4) in the regression equations. Before proceeding to do this, however, it is necessary to examine the data more critically, for it is essential to the strict application of the theory that the assumptions inherent in its development should be satisfied by the data.

(*b*) *Tests of Normality and Linearity*

The primary assumption of the development is that the variables are random samples from bivariate or multivariate normal distributions, and this implies marginal normal distributions, linear regressions, and uniform array variability. From inspection of the plotted data (Figs. 3-6) one would infer that these conditions are not seriously violated. Statistical tests of normality and linearity have been restricted to the control treatment, but there is little reason to believe that these results will not be representative of the other treatments as well. For the significance of departure from normality, the usual tests of asymmetry (Kendall 1946) and kurtosis (Geary and Pearson 1938) have been employed and the results are set out in Table 1.

The thirteen variables listed in Table 1 are not all independent; in fact, the maximum number of independent values in a set is five. In only one case, W_{23e} (logs), is the asymmetry significant but there is a suggestion of positive asymmetry (n.s.) in the absolute values and negative asymmetry ($P<0.05$) in the logarithmic values. This applies whether one considers the five distributions of weight or the corresponding leaf area distributions. The only individual

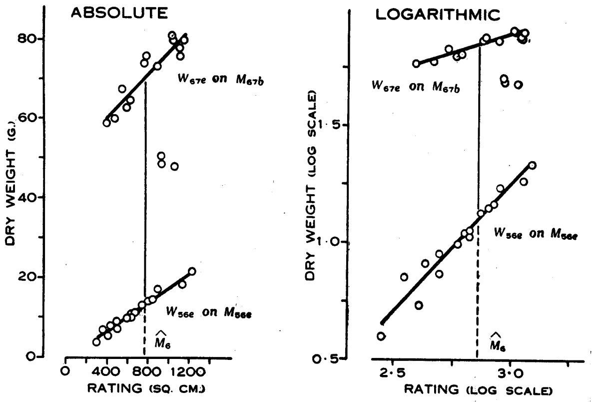

Fig. 6.—The relations between dry weight of plants (tops only) at harvests 6 and 7 and their ratings (leaf areas) at the commencement of harvest interval 6-7. The pairs of regression lines are linked by the maximum likelihood estimates, $\hat{M}_6$ of the true rating as ordinate. The three aberrant values for harvest 7 are not included in the regressions for that harvest (see text).

departures of kurtosis from expectation which are exceptional are associated with the thirteen plants of interval 6-7. Taken over sets of independent distributions the mean departures are not significant.

Table 1

TESTS OF NORMALITY

	Absolute				Logarithmic				
Variable	n	Asymmetry $t=k_3/k_2{}^{3/2}$	σ_t	Kurtosis $a=\frac{M.D.}{S.D.}$	Variable	n	Asymmetry $t=k_3/k_2{}^{3/2}$	σ_t	Kurtosis $a=\frac{M.D.}{S.D.}$
W_{23e}	16	0.7621	0.5643	0.7672	W_{23e}	16	−1.1047	0.5643	0.7625
W_{34e}	16	0.3277	0.5643	0.8540	W_{34e}	16	−0.5305	0.5643	0.8638
W_{45e}	16	0.2359	0.5643	0.8303	W_{45e}	16	−0.4337	0.5643	0.7995
W_{56e}	16	0.5280	0.5643	0.8152	W_{56e}	16	−0.4481	0.5643	0.7999
W_{67e}	13	−0.4917	0.5979	0.8962	W_{67e}	13	−0.5645	0.5979	0.8949
LA_{23e}	16	0.6451	0.5643	0.8502	LA_{23e}	16	−0.7769	0.5643	0.7661
LA_{34b}	16	0.4065	0.5643	0.8450	LA_{34b}	16	−0.3856	0.5643	0.8763
LA_{43e}	16	0.1747	0.5643	0.8581	LA_{34e}	16	−0.4070	0.5643	0.8821
LA_{45b}	16	0.9555	0.5643	0.8011	LA_{45b}	16	−0.0318	0.5643	0.8075
LA_{45e}	16	0.4446	0.5643	0.8535	LA_{45e}	16	−0.2607	0.5643	0.8358
LA_{56b}	16	0.6837	0.5643	0.7905	LA_{56b}	16	−0.4153	0.5643	0.7960
LA_{56e}	16	0.6848	0.5643	0.7945	LA_{56e}	16	−0.1556	0.5643	0.8186
LA_{67b}	13	−0.1153	0.5979	0.8953	LA_{67b}	13	−0.4967	0.5979	0.8660

In the test of linearity (Table 2) all pairs of variables which have been examined are recorded. There is no reason to believe that they constitute a selection which is in any sense biassed. In general, the quadratic term is not

TABLE 2
TESTS OF LINEARITY

Variables	Absolute				Logarithmic			
	Linear Regression Coefficient	Linear Residual Variance	Variance due to Quadratic Term	Correlation Coefficient	Linear Regression Coefficient	Linear Residual Variance	Variance due to Quadratic Term	Correlation Coefficient
LA_{34e} on LA_{34b}	2.6256	189	454	0.9776	0.9180	0.00292	0.00012	0.9744
LA_{45e} on LA_{45b}	1.6569	2823	7211	0.9445	0.7785	0.00369	0.00113	0.9457
LA_{56e} on LA_{56b}	2.0334	7074	82	0.9531	0.8387	0.00427	0.00676	0.9353
W_{23e} on LA_{23e}	0.0126	0.00448	0.00034	0.9819	1.1687	0.00484	0.01045	0.9817
W_{34e} on LA_{34b}	0.0412	0.03757	0.02170	0.9818	1.0378	0.00229	0.00196	0.9840
W_{45e} on LA_{45e}	0.0181	0.40406	0.10426	0.9773	1.0364	0.00222	0.00035	0.9716
W_{56e} on LA_{56b}	0.0382	0.56633	0.56094	0.9888	0.9828	0.00110	0.00002	0.9868
W_{67e} on LA_{67b}	0.0278	9.50690	29.30765	0.9288	0.2979	0.00029	0.00001	0.9441
LW_{23e} on LA_{23e}	0.0088	0.00214	0.00007	0.9824	1.1555	0.00468	0.01329	0.9819
LW_{34e} on LA_{34b}	0.0281	0.01614	0.00806	0.9831	1.0240	0.00208	0.00133	0.9851
LW_{45e} on LA_{45e}	0.0104	0.15012	0.04040	0.9746	0.9470	0.00206	0.00033	0.9686
LW_{56e} on LA_{56b}	0.0202	0.41616	1.13903	0.9714	0.9225	0.00162	0.00002	0.9782
LW_{67e} on LA_{67b}	0.0140	2.61052	3.29875	0.9230	0.5071	0.00120	0.00000	0.9241

significant for either the absolute or logarithmic values. Because of this, the residual variance after removing only the variance due to the linear term has been given.

Homogeneity of array variance has not been examined statistically but from inspection it appears that for logarithmic data the arrays associated with low ratings are more variable than those associated with high ratings. The converse may apply to the absolute data.

Summarizing, the variables are highly correlated with no evidence of departure from a linear trend. There is a tendency to negative asymmetry in the marginal distributions for logarithmic values and possibly greater array variability associated with lower ratings. It is certain that, under conditions of linearity and high correlation, bias introduced from small departure from normality in the marginal distributions will be negligible, so that there is little objection to the application of normal distribution theory to this data. Preferably one would consider only the absolute values but for purposes of illustration and contrast the logarithmic values have also been used.

Several points of interest emerge from an examination of the bivariate distributions of Figures 3-6. The regressions of weight (or leaf weight) on leaf area of samples at two successive harvests referred to leaf area at the first harvest tend to diverge markedly for the absolute data, the regression coefficients often being as much as three times as great at the end as at the beginning of the interval. With the logarithmic data, however, the tendency is reversed, though the regressions are much more nearly parallel (e.g. Figs. 3 and 5). Then, too, the residual variance is much greater for the second than for the first harvest of each pair on the absolute basis, and it tends to be rather less on the logarithmic basis. These trends are confirmed for the other treatments and must be regarded as real for the stages of growth covered by the experiment. In consequence, the much simpler variance expressions which would result on the basis of parallel slopes and equal array variance are not permissible here (see Goodall 1945).

The data for harvest interval 6-7 (Fig. 6) differ from the rest in that the interval was fourteen instead of seven days. All lateral shoots had been nipped out at an early stage up to the time of harvest 6, but this procedure was neglected during a period of wet weather just after this harvest. In consequence, there was a "flush" growth of upper laterals which could not be removed because of their contribution to the weight increment. It is probable that this "flush" growth was more pronounced in the smaller plants within each treatment, and that this helped to bring about the near-parallelism of the regressions for the absolute data and the very pronounced convergence of the regressions for the logarithmic data.

Finally, it will be noted that three values for harvest 7 (Fig. 6) are far short of the weights predicted by the regression for the remaining thirteen values. It is believed, though there are no specific records to confirm it, that these discrepancies are due to a genetically controlled character causing "blindness" of the apical meristem. A number of plants, irrespective of treatment, had shown this condition and had been rejected accordingly. However, in the presence of the "flush" growth mentioned above, it is likely that the fault was overlooked in the three plants in question. In all, six such aberrant values were detected for harvest 7 of the experiment. The phenomenon is of interest in itself, but its detection in this way points to the value of the rating technique for the detection and, where justifiable, the rejection of such aberrant values.

(c) *Calculation of Maximum Likelihood Estimates of the Mean Ratings*

The expression for the likelihood (equation 1) was extended to include additional links in the chain, and the normal equations were formed by differentiating with respect to the four required mean parameters, M_3, M_4, M_5, and M_6. These equations were solved using the reciprocal matrix method. The coefficients of the separate sample means contributing to the estimates of mean ratings at the four harvests (see Table 3) satisfy the requirements for unbiassed estimates.

In Table 4 are given the separate mean ratings of the same harvest date, the means of all ratings at this date (means of 32), and the maximum likelihood estimate. As might be expected, the variance of the mean of all ratings on a given harvest date is approximately half of those of the separate mean ratings. The maximum likelihood estimate, which takes into account the linkage with the ratings of the remaining three harvests, carries with it a further reduction in the variance. The relation of the maximum likelihood estimate to the ratings of the separate samples is indicated in Figure 2, and the improvement of the estimates of the means is evident from their smooth time trends.

TABLE 3

COEFFICIENTS OF THE SEPARATE SAMPLE MEANS CONTRIBUTING TO THE MAXIMUM LIKELIHOOD ESTIMATES OF THE MEAN RATINGS

Sample Mean	Absolute $\hat{M}_3$	$\hat{M}_4$	$\hat{M}_5$	$\hat{M}_6$	Logarithmic $\hat{M}_3$	$\hat{M}_4$	$\hat{M}_5$	$\hat{M}_6$
$\overline{M}_{23e}$	0.2474	0.6088	0.7615	1.3997	0.1793	0.1457	0.0961	0.0683
$\overline{M}_{34b}$	0.7526	−0.6088	−0.7615	−1.3997	0.8207	−0.1457	−0.0961	−0.0683
$\overline{M}_{34e}$	−0.1652	0.5305	0.6635	1.2196	−0.6112	0.3884	0.2563	0.1821
$\overline{M}_{45b}$	0.1652	0.4695	−0.6635	−1.2196	0.6112	0.6116	−0.2563	−0.1821
$\overline{M}_{45e}$	−0.0668	−0.1899	0.5178	0.9518	−0.4421	−0.4424	0.5557	0.3949
$\overline{M}_{56b}$	0.0668	0.1899	0.4822	−0.9518	0.4421	0.4424	0.4443	−0.3949
$\overline{M}_{56e}$	−0.0160	−0.0456	−0.1157	0.6912	−0.2721	−0.2722	−0.2734	0.6530
$\overline{M}_{67e}$	0.0160	0.0456	0.1157	0.3088	0.2721	0.2722	0.2734	0.3470

(*d*) *Improved Estimates of Total Weight, Leaf Weight, and Leaf Area*

By using the additional information available on the population mean rating at each harvest it is now possible to make improved estimates of total weight, leaf weight, and leaf area. This is done, as already indicated for total weight, by substituting the estimates of the mean ratings in the regression equations of Section (*a*) above. These estimates of weight are given in Table 5 and also graphically for maximum likelihood estimates of mean rating in Figures 3-6. In Figure 7, also, all regressions for the logarithmic data are combined in one diagram, and the estimates of weight are shown for the maximum likelihood estimates of mean rating.

The ratio of the variance of the unadjusted to adjusted values cannot exceed the ratio of the total information on the rating to the information from the bivariate distribution alone. The advantage of the extra information in maximum likelihood estimates relative to mean rating at first harvest is here apparent.

The variance of differences of $\hat{W}$ in different treatments is the sum of the corresponding variances. The appropriate variance for differences of $\hat{W}$ from successive harvests of the same treatment is given by substitution in (**9**) or (**10**). For differences of $\hat{W}$ from the same treatment but not from successive harvests one can choose reference points which are not correlated if using mean

TABLE 4

A. RATINGS = LEAF AREA (ABSOLUTE VALUES)

Harvest		Experimental Values					Estimates of True Rating				
		Rating of Harvested Sample			Rating of Next Harvest Sample			Mean of all Ratings at Harvest			Maximum Likelihood Estimate	
		Mean	Variance		Mean	Variance		Mean	Variance		Mean	Variance
3	M_{23e}	43.33	44.51	M_{34b}	45.86	34.57	$\overline{M}_3$	46.09	19.13	$\hat{M}_3$	48.47	11.01
4	M_{34e}	133.33	249.34	M_{45b}	174.63	496.70	$\overline{M}_4$	153.98	194.25	$\hat{M}_4$	140.71	77.01
5	M_{45e}	395.56	1526.31	M_{56b}	291.63	974.13	$\overline{M}_5$	343.59	692.05	$\hat{M}_5$	337.30	244.56
6	M_{56e}	666.56	4509.50	M_{67b}	829.81	3883.65	$\overline{M}_6$	748.18	2245.52	$\hat{M}_6$	766.19	1199.43

B. RATINGS = LEAF AREA (LOGARITHMIC VALUES)

Harvest		Experimental Values					Estimates of True Rating					
		Rating of Harvested Sample			Rating of Next Harvest Sample			Mean of all Ratings at Harvest			Maximum Likelihood Estimate	
		Mean	Variance		Mean	Variance		Mean	Variance		Mean	Variance
3	M_{23e}	1.5850	0.005500	M_{34b}	1.6013	0.003799	$\overline{M}_3$	1.5932	0.002252	$\hat{M}_3$	1.6388	0.000986
4	M_{34e}	2.0715	0.003372	M_{45b}	2.1907	0.003001	$\overline{M}_4$	2.1311	0.001656	$\hat{M}_4$	2.1095	0.000802
5	M_{45e}	2.5637	0.002033	M_{56b}	2.4246	0.002471	$\overline{M}_5$	2.4942	0.001246	$\hat{M}_5$	2.4990	0.000531
6	M_{56e}	2.7906	0.001987	M_{67b}	2.8973	0.001381	$\overline{M}_6$	2.8440	0.000907	$\hat{M}_6$	2.8597	0.000479

N.B.—$\overline{M}_3$ is the mean of $\overline{M}_{23e}$ and $\overline{M}_{34b}$ and so on. The symbol is introduced here as alternative to $\hat{M}$ to simplify presentation.

TABLE [illegible]

A. TOTAL DRY WEIGHT (ABSOLUTE VALUES)

	Experimental Values		Estimated Values					
	Mean	Variance		Mean	Variance		Mean	Variance
W_{23e}	0.5588	0.0073	W_{23e} referred to $\overline{M}_3$	0.5558	0.0038	W_{23e} referred to $\widehat{M}_3$	0.5857	0.0020
W_{34e}	1.8964	0.0609	W_{34e} referred to $\overline{M}_3$	1.9059	0.0317	W_{34e} referred to $\widehat{M}_3$	2.0039	0.0205
			$\overline{M}_4$	2.2103	0.0323	$\widehat{M}_4$	2.0086	0.0203
W_{45e}	7.1056	0.5262	W_{45e} referred to $\overline{M}_4$	6.4716	0.2966	W_{45e} referred to $\widehat{M}_4$	6.0642	0.1104
			$\overline{M}_5$	6.1605	0.2778	$\widehat{M}_5$	6.0461	0.1004
W_{56e}	11.4269	1.4780	W_{56e} referred to $\overline{M}_5$	13.4116	0.7578	W_{56e} referred to $\widehat{M}_5$	13.1710	0.3656
			$\overline{M}_6$	12.8779	0.7695	$\widehat{M}_6$	13.1980	0.3697
W_{67e}	71.5562	4.8807	W_{67e} referred to $\overline{M}_6$	70.1886	2.4606	W_{67e} referred to $\widehat{M}_6$	70.6900	1.7069
B. TOTAL DRY WEIGHT (LOGARITHMIC VALUES)								
W_{23e}	$\bar{1}.6433$	0.007795	W_{23e} referred to $\overline{M}_3$	$\bar{1}.6529$	0.004060	W_{23e} referred to $\widehat{M}_3$	$\bar{1}.7062$	0.001542
W_{34e}	0.2128	0.004226	W_{34e} referred to $\overline{M}_3$	0.2044	0.002190	W_{34e} referred to $\widehat{M}_3$	0.2517	0.001146
			$\overline{M}_4$	0.2777	0.002244	$\widehat{M}_4$	0.2542	0.001160
W_{45e}	0.8144	0.002314	W_{45e} referred to $\overline{M}_4$	0.7644	0.001272	W_{45e} referred to $\widehat{M}_4$	0.7462	0.000679
			$\overline{M}_5$	0.7424	0.001232	$\widehat{M}_5$	0.7473	0.000658
W_{56e}	1.0184	0.002451	W_{56e} referred to $\overline{M}_5$	1.0868	0.001263	W_{56e} referred to $\widehat{M}_5$	1.0915	0.000527
			$\overline{M}_6$	1.0760	0.001304	$\widehat{M}_6$	1.0930	0.000547
W_{67e}	1.8522	0.000191	W_{67e} referred to $\overline{M}_6$	1.8424	0.000092	W_{67e} referred to $\widehat{M}_6$	1.8470	0.000067
C. TOTAL DRY WEIGHT (ANTILOGARITHMIC VALUES)								
W_{23e}	0.4398	0.0080	W_{23e} referred to $\overline{M}_3$	0.4497	0.0043	W_{23e} referred to $\widehat{M}_3$	0.5084	0.0021
W_{34e}	1.6323	0.0597	W_{34e} referred to $\overline{M}_3$	1.6010	0.0298	W_{34e} referred to $\widehat{M}_3$	1.7852	0.0194
			$\overline{M}_4$	1.8954	0.0427	$\widehat{M}_4$	1.7956	0.0198
W_{45e}	6.5223	0.5219	W_{45e} referred to $\overline{M}_4$	5.8130	0.2278	W_{45e} referred to $\widehat{M}_4$	5.5744	0.1119
			$\overline{M}_5$	5.5259	0.1995	$\widehat{M}_5$	5.5886	0.1090
W_{56e}	10.4328	1.4144	W_{56e} referred to $\overline{M}_5$	12.2124	0.9987	W_{56e} referred to $\widehat{M}_5$	12.3452	0.4258
			$\overline{M}_6$	11.9124	0.9811	$\widehat{M}_6$	12.3880	0.4451
W_{67e}	71.1541	5.1270	W_{67e} referred to $\overline{M}_6$	69.5665	2.3606	W_{67e} referred to $\widehat{M}_6$	70.3072	1.7559

ratings at a harvest, so that simple addition of variances will apply. Using maximum likelihood estimates of mean rating the correction to the sum of the variances is of the form

$$-b_k \operatorname{covar}(\bar{y}_a - b_a\bar{x}_a)\hat{x}_k - b_a \operatorname{covar}(\bar{y}_k - b_k\bar{x}_k)\hat{x}_a - 2b_ab_k \operatorname{covar} \hat{x}_k\hat{x}_a.$$

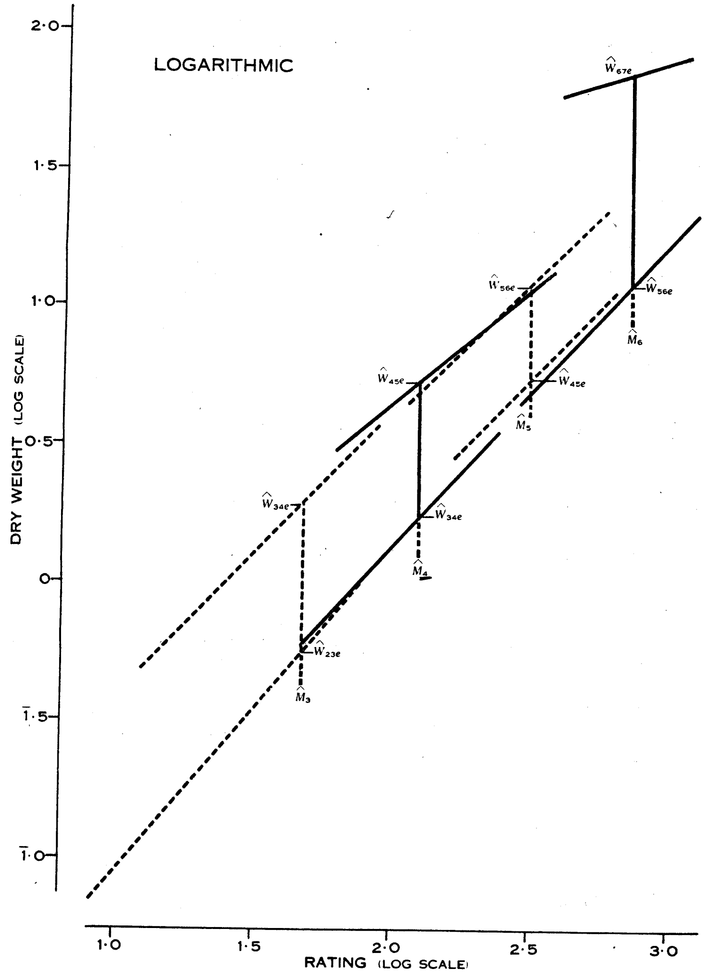

Fig. 7.—The pairs of regression lines (logarithmic series only) of Figs. 3-6 plotted together to illustrate the complete series of linked observations. The first three pairs are for harvest intervals of 7 days, but the last pair of regression lines is for an interval of 14 days.

This expression is dominated by the final term which is always negative so that the sum of the variances is in excess of an unbiassed estimate of the variance of the difference.

The contrast of absolute estimates and the antilogarithms of the logarithmic estimates is essentially parallel to a contrast of arithmetic and geometric means, the latter always being the smaller for distributions of positive values.

(e) *Improved Estimates of* R, E_{LW}, *and* E_{LA}

Direct substitution of the values of $\overline{W}$, $\overline{LW}$, $\overline{LA}$, and $\widehat{W}$, $\widehat{LW}$, $\widehat{LA}$ in the formulae for R, E_{LW}, and E_{LA} and their standard errors, gives the values of Table 6.

Generally speaking the gains in precision in the relative growth rate and net assimilation rates through the use of ratings are quite substantial, principally owing to the fact that the variance of the estimated mean rating is less important since its coefficient involves the differences of regression coefficients or weighted regression coefficients. This also means that there is little advantage in the use of maximum likelihood estimates in place of mean rating at first harvest except where the regression coefficients on the logarithmic basis for W, LW, or LA for the two harvests are very different.

The gains in precision using ratings are indirectly confirmed in this experiment by the much more regular changes in R etc., from one interval to the next in comparison of treatments, R being adjusted independently of course for each treatment.

The principal use of $R(E_{LW}, E_{LA})$ lies in the comparison of treatments over the same period of time and the variance of differences in R between treatments will be the sum of the variances of R for the separate treatments. An unbiassed estimate of the variance of the difference in R of the same treatment but not consecutive intervals is given by the sum of the separate variances for estimates based on mean rating at harvest. For estimates based on maximum likelihood rating the sum of variances would require a correction analogous to the correction for differences in $\widehat{W}$ in the previous section. For differences of R from consecutive intervals the sampling error in the material common to the two intervals will introduce positive corrections to the sum of the separate variances, whether ratings are used or not. For example, using natural logarithms of the weights, the variances of R (see 18) for two successive intervals, without ratings, can be expressed as

$$\frac{1}{(t_2-t_1)^2}\left(\sigma^2_{\overline{U}_{2e}}+\sigma^2_{\overline{U}_{1e}}\right),\frac{1}{(t_3-t_2)^2}\left(\sigma^2_{\overline{U}_{3e}}+\sigma^2_{\overline{U}_{2e}}\right).$$

The correction to the sum of these to give an unbiassed estimate of the differences of R's from the consecutive intervals is

$$\frac{2\sigma^2_{\overline{U}_{2e}}}{(t_2-t_1)(t_3-t_2)}.$$

TABLE 6

RELATIVE GROWTH RATE, R AND NET ASSIMILATION RATES, E_{LW} AND E_{LA}

	Absolute						Logarithmic					
			Rated						Rated			
	Not Rated		Referred to Mean Rating at First Harvest		Referred to Maximum Likelihood Rating		Not Rated		Referred to Mean Rating at First Harvest		Referred to Maximum Likelihood Rating	
Interval	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Relative Growth Rate, R												
3-4	0.1746	0.0287	0.1760	0.0049	0.1757	0.0055	0.1873	0.0361	0.1814	0.0075	0.1794	0.0073
4-5	0.1887	0.0236	0.1535	0.0061	0.1579	0.0083	0.1979	0.0266	0.1601	0.0074	0.1618	0.0077
5-6	0.0679	0.0211	0.1107	0.0038	0.1112	0.0044	0.0671	0.0227	0.1133	0.0038	0.1132	0.0048
6-7	0.1310	0.0079	0.1211	0.0036	0.1199	0.0025	0.1423	0.0085	0.1261	0.0047	0.1240	0.0031
Net Assimilation Rate, E_{LW}												
3-4	0.2499	0.0411	0.2520	0.0084	0.2517	0.0081	0.2670	0.0516	0.2586	0.0105	0.2560	0.0106
4-5	0.2906	0.0375	0.2350	0.0121	0.2408	0.0130	0.3020	0.0422	0.2428	0.0117	0.2448	0.0116
5-6	0.1135	0.0354	0.1865	0.0072	0.1864	0.0072	0.1110	0.0376	0.1876	0.0077	0.1877	0.0077
6-7	0.3429	0.0209	0.3183	0.0121	0.3141	0.0083	0.3616	0.0228	0.3338	0.0147	0.3267	0.0101
Net Assimilation Rate, E_{LA}												
3-4	0.2322	0.0396	0.2342	0.0085	0.2346	0.0083	0.2402	0.0465	0.2325	0.0093	0.2333	0.0095
4-5	0.3086	0.0391	0.2504	0.0132	0.2568	0.0140	0.3190	0.0436	0.2557	0.0116	0.2570	0.0116
5-6	0.1189	0.0364	0.1957	0.0052	0.1957	0.0054	0.1161	0.0395	0.1979	0.0060	0.1980	0.0061

Logarithmic estimates of R, E_{LW}, and E_{LA} tend to be higher than the absolute values in this data while standard errors are of the same order. These relations hold also for other treatments.

VII. General Discussion

In attempting to assess the value of the technique, one must not overlook the computational effort that is involved. There will be obvious cases where random sampling of relatively large numbers of plants (e.g. harvests 1 and 2 of the example) can be accomplished with ease in the time available for harvesting and preparation of the material. In such cases there would be little point in adopting the procedure.

There are also types of experiment, particularly with potted plants, where it is possible to gain statistical control of plant variability by allotting all treatments at random within size groups based on leaf area prior to application of treatments. Any additional gain in precision resulting from the rating technique would rarely justify the computational effort involved unless there was evidence of interaction in the logarithms of yields between treatments and size classes which would be associated with different regression slopes of yield on size for different treatments.

For this event the weights would be referred back to the general mean of the pretreatment ratings and the rating values would be regarded as fixed for repeated sampling. The variance expressions given would then be modified by the omission of reference point variation and the actual deviates of the particular mean ratings from the general mean would be used in conjunction with errors in regression coefficients instead of the expectation of these deviations. Thus, under these conditions (8) would become

$$\frac{\sigma^2_{y_{2e}\cdot x_{2b}}}{n_2} + \frac{\sigma^2_{y_{1e}\cdot x_{1e}}}{n_1} + \sigma^2 b_2(\bar{x}_{2b} - \hat{\bar{x}})^2 + \sigma^2_{b1}(\bar{x}_{1e} - \hat{\bar{x}})^2$$

and this would simplify further on the assumption of the same array variance and regression slope for the two particular distributions to

$$\sigma^2_{y.x}\left(\frac{1}{n_2} + \frac{1}{n_1}\right) + \sigma^2_b(\bar{x}_{2b} - \bar{x}_{1e})^2.$$

In a similar manner the expressions for R, E_{LW}, and E_{LA} would be modified and in general simplified.

The technique outlined in this paper has its chief value in cases where the treatments are operative from germination or from a stage where ratings are unavailable or of little use and where the amount of material at each harvest is severely restricted, for example because of inherent difficulties in harvesting within some suitable unit of time. In growth experiments it is not uncommon for plants to double or even to treble their size each week during early stages of growth, so that the size of the plants soon sets the limit to the number that can be handled in one day. Then, too, the need for precision is perhaps even greater in growth experiments than in yield trials, for the interest centres more in weight increment than in final weight. It is this gain in precision which is so important for experiments on field-grown crops, where plant variation tends to be very great.

Other worthwhile applications of the technique are likely to be in connection with individual leaf studies where high precision is desirable with limited material, and to studies of fruit growth. An experiment has come to our notice in which it was desired to compare fresh and dry weight increments of fruit from trees which had been subjected to varying degrees of thinning. The removal of the necessarily large successive samples of fruit was such as to vitiate the treatments under comparison, and it seems probable that the use of ratings based on an estimate of fruit volume (e.g. the cube of mean diameter) might reduce the necessary sample size to such an extent that treatment would not be seriously affected. That fruit volume is likely to be a good basis for rating in such studies is indicated by the data of Ross (1946) for tomato fruits.

A point which cannot be too strongly emphasized is that the rating must be highly correlated with the yield function for best results. Leaf area is probably the best basis for rating in young plants, but stem measurements (e.g. height $\times$ girth) might be better for mature plants. If photographic standards are used for leaf-area rating, it is highly desirable that the same person should make all the estimates.

In the example of this paper a comparison has been made of the use of the absolute and the logarithmic data. In applying the analysis of variance in the statistical treatment of growth data, it has usually been found necessary to use the logarithmic transformation in order to eliminate the correlation of class means with their standard errors. Furthermore, it was known (Goodall 1945) that the logarithmic data were likely to give approximately parallel regression lines (see also Figs. 3, 4, and 5 of this paper). However, the gain in information using ratings seems to be as great with the untransformed data, so there seems little point in using the transformation, especially as the actual measurements were found to be slightly more compatible with the assumptions underlying the development of the theory.

The computational effort concerned with the determination of the maximum likelihood estimates of the mean ratings could be eliminated in some types of experiment. Thus, if the time interval between the first and the last of the harvests is short, the weight-rating correlations will all be high, and it may be possible to rate all the plants of the experiment at one time. In such a case the mean rating of each treatment automatically becomes the maximum likelihood estimate for the treatment. Even if the weight-rating correlation is not maintained, or it is not possible to rate all the plants at once, the same principle could be extended to two or more groups of harvests. By ignoring the rating linkages between such groups, one would lose a little information, but this is likely to be small in comparison with the computational labour of recovering this information.

VIII. Acknowledgment

The work described in this paper was carried out as part of the research programme of the Division of Plant Industry, C.S.I.R.O.

IX. References

Geary, R. C., and Pearson, E. S. (1938).—"Tests of Normality." (Biometrika Office: London.)

Goodall, D. W. (1945).—*Ann. Bot.* N.S. **9**: **101.**

Kendall, M. G. (1946).—"The Advanced Theory of Statistics." Vol. II, p. 105. (Charles Griffin: London.)

Ross, A. A. (1946).—*Qd. J. Agric. Res.* 3: 121.