

KEY-FACTOR/KEY-STAGE ANALYSIS FOR LIFE TABLE DATA

KOJI YAMAMURA¹

Laboratory of Population Ecology, National Institute of Agro-Environmental Sciences,
Tsukuba, Ibaraki 305-8604, Japan

Abstract. Organisms are subjected to various mortality factors, each of which causes different mortality in each life stage of the organism. The factor that contributes most to the population fluctuation should be called the “key factor.” Similarly, the stage that contributes most to the population fluctuation should be called the “key stage.” The conventional key-factor analyses divide the variation of the total mortality through all stages, K , into the variation of the mortality in the i th stage, k_i . These analyses identify the key stage but do not identify the key factor, unless each stage is subjected to only a single mortality factor. Hence, the conventional key-factor analyses should be called key-stage analyses. A “key-factor/key-stage analysis” is proposed by integrating the conventional key-factor analyses and ANOVA, emphasizing the importance of discriminating between the key factor and the key stage. This analysis identifies the key factor, the key stage, and the combination of factor and stage that is most influential in determining the fluctuation of total mortality.

Key words: division of variance; fluctuation of mortality; key-factor/key-stage analysis; life stage; life table; mortality factor; population fluctuation.

INTRODUCTION

Organisms are subjected to various mortality factors, such as natural enemies and climate, each of which causes different mortality in each life stage of the organism. Some factors may affect the mortality of both the egg stage and the larval stage, while other factors may affect only the larval stage. The factor that contributes most to the population fluctuation should be called the “key factor” (Morris 1959). Similarly, the stage that contributes most to the population fluctuation should be called the “key stage.” To learn why population density varies in place as well as in time, life tables should be compiled simultaneously for a number of different environments. These data could be used to show which factor is the key factor, which stage is the key stage, and through which stage the key factor exerts its greatest effect.

It generally is laborious to examine various environmental factors simultaneously. For this reason, Morris (1959) proposed a “single-factor analysis.” In this analysis, only the mortality caused by a single factor, which is suspected to be a key factor, is analyzed. If the variability of mortality caused by this factor is sufficiently large, it can be confirmed that the factor is a key factor. Morris applied this analysis to the percentage of parasitism for a black-headed budworm, *Acleris variana* (Fern.). However, the mortality caused by a specific factor is not always measurable. In many cases, the mortality of each life stage can only be estimated, since the mortality may be caused by a mixture of

several factors. If the single-factor analysis were applied on a stage-by-stage basis, the analysis would identify the key stage, but would not identify the key factor, unless the stage is subjected to only a single mortality factor. Subsequent to Morris’s work, several authors improved the single-factor analysis under the terminology of “key-factor analysis” (Varley and Gradwell 1960, Mott 1966, Metcalfe 1972, Smith 1973, Podoler and Rogers 1975, Manly 1977). In these studies, however, factors and stages were confused, and this confusion seems to remain today.

In this paper, I emphasize the importance of discriminating between the key factor and the key stage. First, I discuss the conventional key-factor analyses that should be called “key-stage analyses.” Second, I discuss the utilization of ANOVA as a key-factor analysis. Third, I propose a “key-factor/key-stage analysis” by integrating the conventional key-factor analysis and ANOVA. The effectiveness of the analysis is demonstrated by using the life table data of *Pieris rapae crucivora* Boisduval (Lepidoptera, Pieridae).

KEY STAGE ANALYSIS

The conventional key-factor analyses will be summarized as follows. Let N_{ij} be the population entering the i th life stage of the j th observation. Let k_{ij} be the mortality of the i th life stage of the j th observation, defined as the negative logarithm of the survival rate: $k_{ij} = -\log(N_{i+1,j}/N_{ij})$, ($i = 1, 2, \dots, s$; $j = 1, 2, \dots, n$). Let K_j be the mortality through all stages, i.e., $K_j = \sum_{i=1}^s k_{ij}$. Here, “mortality” will be used in a broad sense to cover any loss in a given population, whether this loss results from direct mortality, from dispersal, or from reduced fecundity (Morris 1957). The mean of

Manuscript received 11 August 1997; accepted 21 April 1998; final version received 12 May 1998.

¹ E-mail: yamamura@niaes.affrc.go.jp

k_{ij} is denoted by $\bar{k}_i = (1/n) \sum_{j=1}^n k_{ij}$; the mean of K_j is denoted by $\bar{K} = (1/n) \sum_{j=1}^n K_j$. Also, the variance of K_j is denoted by $V(K) = [1/(n-1)] \sum_{j=1}^n (K_j - \bar{K})^2$, the variance of k_{ij} by $V(k_i) = [1/(n-1)] \sum_{j=1}^n (k_{ij} - \bar{k}_i)^2$, and the covariance between k_{ij} and k_{hj} by $\text{cov}(k_i, k_h) = [1/(n-1)] \sum_{j=1}^n (k_{ij} - \bar{k}_i)(k_{hj} - \bar{k}_h)$. The relative contribution of the i th-stage mortality to the fluctuation of the total mortality can be evaluated by comparing the deviation of the i th-stage mortality, $(k_{ij} - \bar{k}_i)$, and the deviation of the total mortality, $(K_j - \bar{K})$. Varley and Gradwell (1960) plotted the deviations of mortality on a graph and visually compared the synchronization of the fluctuation.

Several other methods enabling a quantitative comparison are based on the division of the variance into components. Mott (1966) divided $V(K)$ into $V(k_i)$ and $\text{cov}(k_i, k_h)$:

$$V(K) = \sum_{i=1}^s V(k_i) + \sum_{i=1}^s \sum_{g=1}^{s(i \neq g)} \text{cov}(k_i, k_g). \quad (1)$$

This division elucidates the influence of the interaction between the mortality of different stages, as well as the influence of each stage. However, the interpretation of Eq. 1 is not straightforward, since the effect of each stage is scattered among many terms. Smith (1973) and Harcourt (1986) bundled the terms of Eq. 1 symmetrically among stages:

$$V(K) = \sum_{i=1}^s \left[V(k_i) + \sum_{g=1}^{s(i \neq g)} \text{cov}(k_i, k_g) \right] \quad (2)$$

which can be expressed in a simpler form given by

$$V(K) = \sum_{i=1}^s \text{cov}(k_i, K). \quad (3)$$

Podoler and Rogers (1975) used the regression coefficient of k_{ij} against K_j to evaluate the influence of k_{ij} on K_j . Their method is identical to the division used by Smith (1973), since the regression coefficients are proportional to $\text{cov}(k_i, K)$. These kinds of analyses divide the variance into stages, but do not divide it into factors, unless the stage is subjected to only a single mortality factor. Therefore, I call these analyses key-stage analyses instead of key-factor analyses.

KEY-FACTOR ANALYSIS

The deviation of the total mortality, $(K_j - \bar{K})$, which is influenced by the mortality of each stage, is also influenced by various environmental factors, e.g., rainfall, temperature, food quality, and natural enemies. As an illustration, imagine the following field conditions of four plots ($j = 1, 2, 3, 4$): $j = 1$, higher rainfall and higher temperature; $j = 2$, higher rainfall and lower temperature; $j = 3$, lower rainfall and higher temperature; $j = 4$, lower rainfall and lower temperature. Assume that $(K_j - \bar{K})$ increases by the increment B_1 when the rainfall is higher and that $(K_j - \bar{K})$ increases by the increment B_2 when the temperature is higher.

Knowledge of the effects of a variation of the environmental factors on the variation of K_j is desired. Hence, the sum of the effects of each environmental factor should be defined as zero. Furthermore, assume that $(K_j - \bar{K})$ decreases by the decrement B_1 when the rainfall is lower and that $(K_j - \bar{K})$ decreases by the decrement B_2 when the temperature is lower.

The deviation of the total mortality, $(K_j - \bar{K})$, will be further subjected to variabilities that cannot be explained by known factors. Hence, $(K_j - \bar{K})$ should be written as follows:

$$\begin{aligned} K_1 - \bar{K} &= B_1 + B_2 + U_1 \\ K_2 - \bar{K} &= B_1 - B_2 + U_2 \\ K_3 - \bar{K} &= -B_1 + B_2 + U_3 \\ K_4 - \bar{K} &= -B_1 - B_2 + U_4 \end{aligned} \quad (4)$$

where U_j are the unknown variabilities. It is convenient to express these equations by a matrix form:

$$\begin{pmatrix} K_1 - \bar{K} \\ K_2 - \bar{K} \\ K_3 - \bar{K} \\ K_4 - \bar{K} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}. \quad (5)$$

Generally, the deviation $(K_j - \bar{K})$ can be expressed by

$$\mathbf{K} - \bar{\mathbf{K}} = \mathbf{X}\mathbf{B} + \mathbf{U} \quad (6)$$

where \mathbf{K} is a column vector with elements K_j , $\bar{\mathbf{K}}$ is a column vector with elements \bar{K} , \mathbf{X} is a design matrix determining the arrangement of environmental factors, \mathbf{B} is a column vector expressing the effects of factors, and \mathbf{U} is a column vector with elements U_j . It is being assumed that the number of observations is the same for all combinations of the levels of factors.

To identify the key factor, I divide the variance, $V(K) = 1/(n-1) \sum_{j=1}^n (K_j - \bar{K})^2$, into factors. A simple principle to achieve the division is to use the sum of squares in the ANOVA table. In the performance of the ANOVA, treat each factor as an independent variable, and treat \mathbf{K} , (not $\mathbf{K} - \bar{\mathbf{K}}$), as a dependent variable. Then, the column of the sum of squares in the ANOVA table gives a partition of $\sum_{j=1}^n (K_j - \bar{K})^2$. To clarify the relation between the ANOVA and the conventional key-factor analyses in the next section, \mathbf{X}_h is defined as the \mathbf{X} matrix in which only the columns related to the h th factor are included and the other columns are replaced by zeros ($h = 1, 2, \dots, f$). For the example, these matrices are as follows:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 0 \\ -1 & 0 \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

and

TABLE 1. Key-factor, key-stage table for the mortality of larvae of *P. rapae crucivora*.

Factor	df	Instar				Total†
		1st	2nd	3rd	4th	
Block	3	-114	105	272	423	686
Spacing	1	56	-76	172	181	332
Season	1	29	-22	-16	575	565
Spacing × season	1	67	-29	62	13	114
Unknown variability	9	4	-41	361	742	1066
Total‡	15	42	-64	850	1933	2762

Notes: The variance of the total mortality, $V(K)$, is divided into five factors × four stages. All values are multiplied by 10^4 to facilitate the comparison.

† This column shows which factor is the key factor.

‡ This row shows which stage is the key stage.

$$\mathbf{X} = \sum_{h=1}^f \mathbf{X}_h. \tag{7}$$

Then, the sum of squares in the ANOVA table is expressed by $(\mathbf{X}_h \hat{\mathbf{B}})'(\mathbf{X}_h \hat{\mathbf{B}})$, and the residual sum of squares is expressed by $\hat{\mathbf{U}}' \hat{\mathbf{U}}$, where a hat (^) indicates estimation obtained by ANOVA and a prime (') indicates a transposition (see Searle 1971). Hence, $V(K)$ can be expressed as

$$V(K) = \frac{1}{n-1} \sum_{h=1}^f \left[(\mathbf{X}_h \hat{\mathbf{B}})'(\mathbf{X}_h \hat{\mathbf{B}}) \right] + \frac{1}{n-1} [\hat{\mathbf{U}}' \hat{\mathbf{U}}]. \tag{8}$$

Eq. 8 divides the variance into $(f + 1)$ components separated by brackets, each corresponding to a factor and unknown variability, where f is the total number of factors. Hence, the relative contribution of each factor can be evaluated by comparing these $(f + 1)$ terms. This kind of analysis should be called key-factor analysis, since it divides the variance into factors.

KEY-FACTOR/KEY-STAGE ANALYSIS

The key-stage analysis given by Eq. 3 and the key-factor analysis given by Eq. 8 can be combined in the following manner. I have considered the effects of environmental factors on the deviation of the total mortality $(K_j - \bar{K})$. In an actual situation, however, the environmental factors affect each of the life stages. To take this into account, I express the effects of environmental factors on the i th stage by

$$\mathbf{k}_i - \bar{\mathbf{k}}_i = \mathbf{X} \mathbf{b}_i + \mathbf{u}_i \tag{9}$$

where \mathbf{k}_i is a column vector with elements k_{ij} , $\bar{\mathbf{k}}_i$ is a column vector with elements \bar{k}_i , \mathbf{b}_i is a column vector expressing the effects of factors in the i th stage, and \mathbf{u}_i is a column vector expressing the unknown variability in the i th stage. As a logical consequence, $\mathbf{U} = \sum_{i=1}^s \mathbf{u}_i$ and $\mathbf{B} = \sum_{i=1}^s \mathbf{b}_i$. To evaluate the effect of each combination of factor × stage on $V(K)$, the variance and covariance of \mathbf{k}_i in Eq. 2 should be divided into factors. A simple principle to achieve the division is to use the sum of squares and sum of products calculated in MANOVA. The \mathbf{k}_i , (not $\mathbf{k}_i - \bar{\mathbf{k}}_i$), of the different stages are to be treated as different variables in the

performance of MANOVA. Then, the variance is expressible as

$$V(K) = \frac{1}{n-1} \sum_{i=1}^s \sum_{h=1}^f \left[\sum_{m=1}^s (\mathbf{X}_h \hat{\mathbf{b}}_i)'(\mathbf{X}_h \hat{\mathbf{b}}_m) \right] + \frac{1}{n-1} \sum_{i=1}^s \left[\sum_{m=1}^s \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_m \right] \tag{10}$$

where $(\mathbf{X}_h \hat{\mathbf{b}}_i)'(\mathbf{X}_h \hat{\mathbf{b}}_m)$ and $\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_m$ are the elements of the SSP matrix (sum of squares and products matrix) calculated in MANOVA (see Chatfield and Collins 1980). Eq. 10 divides the variance into $s(f + 1)$ components, separated by brackets, each corresponding to a combination of factors × stages. These $s(f + 1)$ terms may be easily obtained by using statistical software such as SAS, JMP, or Systat; $\sum_{m=1}^s (\mathbf{X}_h \hat{\mathbf{b}}_i)'(\mathbf{X}_h \hat{\mathbf{b}}_m)$ is obtained by summing the elements of the i th column of the SSP matrix of the h th factor; and $\sum_{m=1}^s \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_m$ is obtained by summing the elements of the i th column of the error SSP matrix. In the usual MANOVA, the SSP matrix is used to test the effects of factors. However, such kinds of statistical tests are not of interest at present, since the manner in which the total variance is attributed to factors and stages is under investigation.

EXAMPLE

The effectiveness of this method will be demonstrated using the data obtained by the experiment studying the effect of plant spacing on the mortality of *P. rapae crucivora* (Yamamura and Yano, unpublished data). The data are listed in the Appendix so that readers can verify the calculations. In this experiment, cabbage seedlings were planted at two levels of plant spacing: 2×2 m (sparse) and 0.5×0.5 m (dense). The experimental plots, each 10×10 m, were replicated by four blocks, and the number of the first, second, third, fourth, and fifth instar larvae were estimated in each plot. The experiment was repeated twice in 1989, starting on May 25 (spring experiment) and on July 12 (summer experiment).

The result of the analysis is expressed in a “key-factor/key-stage table,” with columns indicating stages and rows indicating factors (Table 1). The “Total” col-

umn lists the sum of squares of each factor given by Eq. 8, showing the degree of influence due to each factor in determining the fluctuation of total mortality. The largest component is the unknown variability (1066) in this example. The "Total" row lists the result of the key-stage analysis given by Eq. 3, showing the degree of influence due to each life stage in determining the fluctuation of total mortality. The mortality of the fourth-instar larvae is most influential (1933) in this example. Furthermore, the key-factor/key-stage table shows the combination of factor \times stage that is influential in determining the total mortality. The largest source of the variance of K is the effect of unknown variability through the mortality of the fourth-instar larvae (742) in this example. The variation in season also significantly increases the variance of K through the mortality of the fourth-instar larvae (575). The effects of block, i.e., the effects of spatial variation, through the mortality of the third- and fourth-instar larvae are also relatively large (272 and 423, respectively).

DISCUSSION

The results of the key-factor/key-stage analysis and those of the conventional key-factor analyses may be sometimes contradictory, since the key-factor/key-stage analysis distinguishes between the cause and the effect. For example, the percentage of parasitism that is directly measured in field is not a factor in the key-factor/key-stage analysis; it is an effect caused by factors such as the number of parasites. As an illustration, consider a situation in which the activity of parasites is highly influenced by the weather condition (Kiritani and Hokyō 1970, for example). If the weather condition fluctuates widely, the percentage of parasitism fluctuates widely. In this case, the conventional key-factor analyses identify the parasitism as the key factor. In contrast, the key-factor/key-stage analysis identifies the weather condition as the key factor and states that the variability of weather condition increases the variance of K through parasitism. Thus, if the environmental factors as well as the mortality of populations could be measured, the key-factor/key-stage analysis could quantitatively identify the ultimate cause of population dynamics.

The key-factor/key-stage analysis can be applied to numerical factors as well as nominal factors if only a single factor is considered. For example, it is possible to evaluate the effect of temperature, given by numerical values instead of nominal levels, on the population variability. In this case, \mathbf{X} in Eq. 6 is a column vector whose elements are the temperature with the mean subtracted. The variance, $V(K)$, can be divided into the effect of temperature and the effect of unknown variability on each stage, using the SSP matrix generated by multivariate linear regression analysis. When more than one factor is analyzed, however, numerical factors will frequently present problems, i.e., if the compo-

nents of design matrix, \mathbf{X}_h , are not mutually orthogonal, the key factor cannot be determined, since $V(K)$ cannot be divided algebraically. This difficulty can be intuitively understood by considering an imaginary situation where the rainfall and the temperature are closely correlated, e.g., more observations are obtained for the combination of high rainfall and high temperature. In such a case, even if the rainfall has a large contribution to the population variability, it cannot be judged whether this contribution derives from the rainfall or from the temperature. Therefore, the experiment should be carefully designed when more than one factor is being considered. It is preferable that the number of observations is the same for all combinations of levels of factors.

It is also possible to analyze the variability of the population that completed the s th stage, $(N_{s+1,j})$, instead of the variability of mortality. For convenience, let k_{0j} be the negative logarithm of the number of individuals entering the first stage, i.e., $k_{0j} = -\log(N_{1j})$. Then, the following relation holds: $\log(N_{s+1,j}) = -\sum_{i=0}^s k_{ij}$. Hence, the variance can be divided into the contribution of each stage by a form similar to Eq. 3:

$$V[\log(N_{s+1})] = \sum_{i=0}^s \text{cov}[k_i, -\log(N_{s+1})]. \quad (11)$$

The variance can be further divided into each combination of factor \times stage by a form similar to Eq. 10:

$$V[\log(N_{s+1})] = \frac{1}{n-1} \sum_{i=0}^s \sum_{h=1}^f \left[\sum_{m=0}^s (\mathbf{X}_h \hat{\mathbf{b}}_i)' (\mathbf{X}_h \hat{\mathbf{b}}_m) \right] + \frac{1}{n-1} \sum_{i=0}^s \left[\sum_{m=0}^s \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_m \right] \quad (12)$$

where the $(f+1)(s+1)^2$ terms seen inside the summation sign are the elements of the SSP matrix calculated in MANOVA in which \mathbf{k}_i of the different stages are treated as different variables.

The estimates of population size are sometimes subjected to large sampling errors. The population data used for the calculation of Table 1 also contain sampling errors; the survival rate exceeds 1 in several elements. If N_{ij} contains such sampling errors, the $\text{cov}(k_i, K)$ in Eq. 3 is overestimated for $i=1$ and $i=s$, and the $\text{cov}[k_i, -\log(N_{s+1})]$ in Eq. 11 is overestimated for $i=s$ (see Kuno 1971). Similarly, the components of the first and final stages are overestimated in Eq. 10, as are components of the final stage in Eq. 12. Thus, the results of the analysis may be biased in this case. Careful identification of the key factor and key stage should be made if the estimates of population sizes contain large sampling errors.

ACKNOWLEDGMENTS

I thank K. Kiritani and K. Iwao for their valuable comments on the manuscript.

LITERATURE CITED

Chatfield, C., and A. J. Collins. 1980. Introduction to Multivariate Analysis. Chapman, London, UK.

Harcourt, D. G. 1986. Population dynamics of the diamondback moth in Southern Ontario. Pages 3–15 in T. D. Griggs, editor. Diamondback moth management: proceedings of the first international workshop. Asian Vegetable Research and Development Center, Shanhua, Taiwan.

Kiritani, K., and N. Hoko. 1970. Studies on the population ecology of the southern stink bug, *Nezara viridula* L. (Heteroptera: Pentatomidae). In Japanese. Ministry of Agriculture, Forestry, and Fisheries, Tokyo, Japan.

Kuno, E. 1971. Sampling error as a misleading artifact in "key factor analysis." *Researches on Population Ecology* **13**:28–45.

Manly, B. F. J. 1977. The determination of key factors from life table data. *Oecologia* **31**:111–117.

Metcalf, J. R. 1972. An analysis of the population dynamics of the Jamaican sugar-cane pest *Saccharosydne saccharivora* (Westw.) (Hom., Delphacidae). *Bulletin of Entomological Research* **62**:73–85.

Morris, R. F. 1957. The interpretation of mortality data in studies on population dynamics. *Canadian Entomologist* **89**:49–69.

———. 1959. Single-factor analysis in population dynamics. *Ecology* **40**:580–588.

Mott, D. G. 1966. The analysis of determination in population systems. Pages 179–194 in K. E. F. Watt, editor. *Systems analysis in ecology*. Academic Press, New York, New York, USA.

Podoler, H., and D. Rogers. 1975. A new method for the identification of key factors from life table data. *Journal of Animal Ecology* **44**:85–114.

Searle, S. R. 1971. *Linear models*. Wiley, New York, New York, USA.

Smith, R. H. 1973. The analysis of intra-generation change in animal populations. *Journal of Animal Ecology* **42**:611–622.

Varley, G. C., and G. R. Gradwell. 1960. Key factors in population studies. *Journal of Animal Ecology* **29**:399–401.

APPENDIX

This appendix shows the data used to calculate Table 1. The population of *P. rapae crucivora* entering each larval instar per plant was estimated under the three factors (plant spacing, season, and block).

Spacing	Season	Block	Instar population				
			1st (N_{1j})	2nd (N_{2j})	3rd (N_{3j})	4th (N_{4j})	5th (N_{5j})
Sparse	Spring	1	15.728	9.226	7.246	6.937	5.911
Sparse	Spring	2	9.370	5.026	3.905	3.451	2.599
Sparse	Spring	3	16.258	9.086	6.464	4.835	3.377
Sparse	Spring	4	10.671	5.390	5.460	5.670	4.079
Sparse	Summer	1	50.427	24.588	22.134	13.212	3.076
Sparse	Summer	2	29.489	13.647	12.844	12.742	5.403
Sparse	Summer	3	32.821	14.863	10.659	9.731	6.638
Sparse	Summer	4	33.043	16.038	15.364	12.559	8.403
Dense	Spring	1	3.977	2.584	2.265	1.972	1.610
Dense	Spring	2	5.039	2.428	2.059	2.019	1.478
Dense	Spring	3	2.777	1.062	0.718	0.689	0.832
Dense	Spring	4	5.666	3.269	2.546	2.705	2.367
Dense	Summer	1	17.281	11.347	9.535	7.911	2.109
Dense	Summer	2	14.522	9.701	7.472	6.106	3.703
Dense	Summer	3	14.056	6.201	3.111	4.629	5.000
Dense	Summer	4	14.240	7.526	7.913	11.283	6.917