APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 1992, p. 3337–3342
0099-2240/92/103337-06$02.00/0

Vol. 58, No. 10

Evaluation of Data Transformations Used with the Square Root and Schoolfield Models for Predicting Bacterial Growth Rate†

SUSAN A. ALBER AND DONALD W. SCHAFFNER*

Department of Food Science, P.O. Box 231, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903

Received 12 May 1992/Accepted 30 July 1992

A comparison was made between mathematical variations of the square root and Schoolfield models for predicting growth rate as a function of temperature. The statistical consequences of square root and natural logarithm transformations of growth rate used in several variations of the Schoolfield and square root models were examined. Growth rate variances of *Yersinia enterocolitica* in brain heart infusion broth increased as a function of temperature. The ability of the two data transformations to correct for the heterogeneity of variance was evaluated. A natural logarithm transformation of growth rate was more effective than a square root transformation at correcting for the heterogeneity of variance. The square root model was more accurate than the Schoolfield model when both models used natural logarithm transformation.

The square root and Schoolfield models are frequently compared for their ability to predict bacterial growth rate or lag time as a function of temperature. Several variations which involve the transformation of growth rate or lag time to either a square root or logarithmic scale have been developed for both models. The original formulation of the square root model utilizes a square root transformation. A logarithm transformation is frequently used with the Schoolfield model (7). The statistical appropriateness of these two transformations of growth rate and the consequences of weighted regression were investigated.

Square root model variates. The square root model, proposed by Ratkwosky et al. (6), is given as:

$$\sqrt{k} = b(T - T_{min})\{1 - \exp[c(T - T_{max})]\} \quad (1)$$

where k is the growth rate (time^{-1}), b is a regression coefficient, T is the temperature (K), T_{min} is the notional minimum growth temperature (K), c is a regression coefficient, and T_{max} is the notional maximum growth temperature (K).

In equation 1, extrapolations above T_{max} result in positive growth rate predictions. A modification of the model which eliminates this effect was proposed by Zwietering et al. (11). The three square root model variates investigated in this study are based on this modified model. The model variate as proposed by Zwietering et al. is:

$$k = [b(T - T_{min})]^2\{1 - \exp[c(T - T_{max})]\} \quad (2)$$

A square root transformation of equation 2 gives:

$$\sqrt{k} = b(T - T_{min})\{1 - \exp[c(T - T_{max})]\}^{1/2} \quad (3)$$

A natural logarithm transformation of equation 2 gives:

$$\ln(k) = \ln([b(T - T_{min})]^2\{1 - \exp[c(T - T_{max})]\}) \quad (4)$$

Schoolfield model variates. The Schoolfield model is derived from Arrhenius rate kinetics. The model was originally proposed by Sharpe and DeMichele (9). The model is based on several simplifying assumptions: the growth rate of an organism at a given temperature is assumed to be governed by a single rate-controlling enzyme which is reversibly inactivated at low and high temperatures; the total concentration of the rate-controlling enzyme, in both the active and inactive states, is assumed to remain constant and independent of temperature; and the growth rate is a function of the ratio of active enzyme to inactive enzyme (8).

The Schoolfield equation is given as:

$$k = \frac{\rho_{(25°C)}\frac{T}{298}\exp\left[\frac{\Delta H^{\neq}_A}{R}\left(\frac{1}{298} - \frac{1}{T}\right)\right]}{1 + \exp\left[\frac{\Delta H_L}{R}\left\{\frac{1}{T_{1/2L}} - \frac{1}{T}\right\}\right] + \exp\left[\frac{\Delta H_H}{R}\left\{\frac{1}{T_{1/2H}} - \frac{1}{T}\right\}\right]} \quad (5)$$

where k is the growth rate (time^{-1}), $\rho_{(25°C)}$ is the growth rate at 25°C (time^{-1}), T is the temperature (K), R is the universal gas constant (8.314 J K^{-1} mol^{-1}), $\Delta H^{\neq}_A$ is the enthalpy of activation of the reaction catalyzed by the rate-controlling enzyme (J mol^{-1}), ΔH_L is the change in enthalpy associated with low-temperature inactivation of the enzyme (J mole^{-1}), $T_{1/2L}$ is the temperature at which the enzyme is 50% inactive because of low temperature (K), ΔH_H is the change in enthalpy associated with high-temperature inactivation of the enzyme (J mol^{-1}), and $T_{1/2H}$ is the temperature at which the enzyme is 50% inactive because of high temperature (K). A square root transformation of equation 5 gives:

$$\sqrt{k} = \sqrt{\frac{\rho_{(25°C)}\frac{T}{298}\exp\left[\frac{\Delta H^{\neq}_A}{R}\left(\frac{1}{298} - \frac{1}{T}\right)\right]}{1 + \exp\left[\frac{\Delta H_L}{R}\left\{\frac{1}{T_{1/2L}} - \frac{1}{T}\right\}\right] + \exp\left[\frac{\Delta H_H}{R}\left\{\frac{1}{T_{1/2H}} - \frac{1}{T}\right\}\right]}} \quad (6)$$

A natural logarithm transformation of equation 5 gives:

$$\ln(k) = \ln\left[\frac{\rho_{(25°C)}\frac{T}{298}\exp\left[\frac{\Delta H^{\neq}_A}{R}\left(\frac{1}{298} - \frac{1}{T}\right)\right]}{1 + \exp\left[\frac{\Delta H_L}{R}\left\{\frac{1}{T_{1/2L}} - \frac{1}{T}\right\}\right] + \exp\left[\frac{\Delta H_H}{R}\left\{\frac{1}{T_{1/2H}} - \frac{1}{T}\right\}\right]}\right] \quad (7)$$

* Corresponding author. Electronic mail address: SCHAFFNER@ZODIAC.RUTGERS.EDU.

† Manuscript D-10122-1-92 of the New Jersey Agricultural Experiment Station.

Statistical considerations. Regression analysis relies on several stochastic assumptions. All observations are assumed to have equal population variances. Homogeneity of variance is important for obtaining an accurate regression line. A second assumption is that each observation comes from a normally distributed population. This assumption is important when making inferences about populations. The construction of accurate confidence intervals is dependent on normally distributed populations.

Ratkowsky et al. called attention to the fact that *Escherichia coli* generation time data from Smith, cited by Ratkowsky et al., violates the assumption of variance homogeneity (7). A weighted least-squares regression analysis can be used to compensate for heterogeneity of variance. An alternative approach which may correct for variance heterogeneity is a data transformation which changes the variance and shape of the population distributions associated with each datum point. A suitable transformation is one which results in equivalent variances. The variances of transformed data can be approximated by the following relationship (2):

$$(\sigma'_i)^2 = \left[\frac{d}{dy_i}[f(y_i)]\right]^2 \sigma_i^2 \tag{8}$$

where $(\sigma'_i)^2$ is the variance of the i^{th} transformed observation, $f(y_i)$ is the transformation applied to y_i, y_i is the i^{th} observation, and σ_i^2 is the variance of y_i.

The variance of the natural logarithm of the i^{th} k $\{\mathrm{Var}[\ln(k_i)]\}$ can be estimated from the variance of the i^{th} k $[\mathrm{Var}(k_i)]$ according to the following formula:

$$\begin{aligned}\mathrm{Var}[\ln(k_i)] &= \left[\frac{d}{dk_i}[\ln(k_i)]\right]^2 \mathrm{Var}(k_i)\\ &= \left[\frac{1}{k_i}\right]^2 \mathrm{Var}(k_i)\\ &= \frac{\mathrm{Var}(k_i)}{k_i^2}\end{aligned} \tag{9}$$

where k_i is the i^{th} growth rate. The variance of the square root of the i^{th} k $[\mathrm{Var}(\sqrt{k_i})]$ can be estimated from $\mathrm{Var}(k_i)$ according to the following formula:

$$\begin{aligned}\mathrm{Var}(\sqrt{k_i}) &= \left[\frac{d}{dk_i}(\sqrt{k_i})\right]^2 \mathrm{Var}(k_i)\\ &= \left[\frac{1}{2\sqrt{k_i}}\right]^2 \mathrm{Var}(k_i)\\ &= \frac{\mathrm{Var}(k_i)}{4k_i}\end{aligned} \tag{10}$$

Transformation of growth rates causes a weighting effect on regression by disproportionally altering the magnitude of each residual. Approximately the same parameter estimates resulting from regression on a transformed model can be produced with the untransformed model variate and an appropriate weighting of the data. A weighting scheme which will produce the same parameter estimates as a natural logarithm transformation can be derived as follows:

$$\text{Weighted SSE} = \sum\left[\frac{1}{\mathrm{Var}(k_i)}(k_i - \hat{k}_i)^2\right] \tag{11}$$

where SSE is the sum of squares due to error, $1/\mathrm{Var}(k_i)$ is the i^{th} weight, k_i is the i^{th} observed growth rate, and $\hat{k}_i$ is the i^{th} predicted growth rate. When $\mathrm{Var}(k_i)$ are all equal, the weights $[1/\mathrm{Var}(k_i)]$ are constant and will not affect parameter estimation. $\mathrm{Var}(k_i)$ in equation 11 can be replaced with $\mathrm{Var}[\ln(k)]k^2$ from equation 9 to give:

$$\text{Weighted SSE} = \sum\left[\frac{1}{\mathrm{Var}[\ln(k_i)]k_i^2}(k_i - \hat{k}_i)^2\right] \tag{12}$$

In an unweighted regression of a natural logarithm-transformed model, $\mathrm{Var}[\ln(k_i)]$ is assumed to be constant. Therefore, it can be removed from equation 12 to give:

$$\text{Weighted SSE} = \sum\left[\frac{1}{k_i^2}(k_i - \hat{k}_i)^2\right] \tag{13}$$

This weighting scheme yields approximately the same parameter estimates when applied to an untransformed model as regression with an unweighted, natural logarithm-transformed model.

A weighting scheme which will produce approximately the same results as a square root transformation by using untransformed data can be derived in a similar fashion as follows:

$$\text{Weighted SSE} = \sum\left[\frac{1}{\mathrm{Var}(k_i)}(k_i - \hat{k}_i)^2\right] \tag{14}$$

Replace $\mathrm{Var}(k_i)$ with $4\mathrm{Var}(\sqrt{k_i})\,k_i$ from equation 10:

$$\text{Weighted SSE} = \sum\left[\frac{1}{4\mathrm{Var}(\sqrt{k_i})\,k_i}(k_i - \hat{k}_i)^2\right] \tag{15}$$

Remove $4\mathrm{Var}(\sqrt{k_i})$:

$$\text{Weighted SSE} = \sum\left[\frac{1}{(k_i)}(k_i - \hat{k}_i)^2\right] \tag{16}$$

This weighting scheme yields approximately the same parameter estimates when applied to an untransformed model as regression using an unweighted, square root-transformed model.

MATERIALS AND METHODS

Strain. An isolate of *Yersinia enterocolitica* serotype 08 was generously provided by Mehdi Shayegani, New York Department of Health. Stock cultures of *Y. enterocolitica* were maintained on brain heart infusion agar (Difco, Detroit, Mich.) slants and transferred monthly.

Experimental procedure. The growth rate of *Y. enterocolitica* in brain heart infusion broth (Difco) was experimentally determined at 15 different temperatures, from 272 to 316 K. The following procedure was performed seperately for each temperature. *Y. enterocolitica* was transferred from a stock culture to 5 ml of brain heart infusion broth, incubated at 301 K, and grown to stationary phase (approximately 24 h). This culture, which had an approximate concentration of 10^9 CFU/ml, was used as the inoculum. The inoculum was then diluted with brain heart infusion broth, and 0.1-ml aliquots of a 10^{-2} dilution were delivered to a series (between 10 and 20) of tubes containing 4.9 ml of brain heart infusion broth, which had been preincubated at the experimental temperature. The inoculated cultures contained approximately $2.0 \times 10^{5.5}$ CFU/ml at time zero.

The inoculated cultures were incubated in a low-temperature incubator (model 146A; Fisher, Pittsburgh, Penn.) at

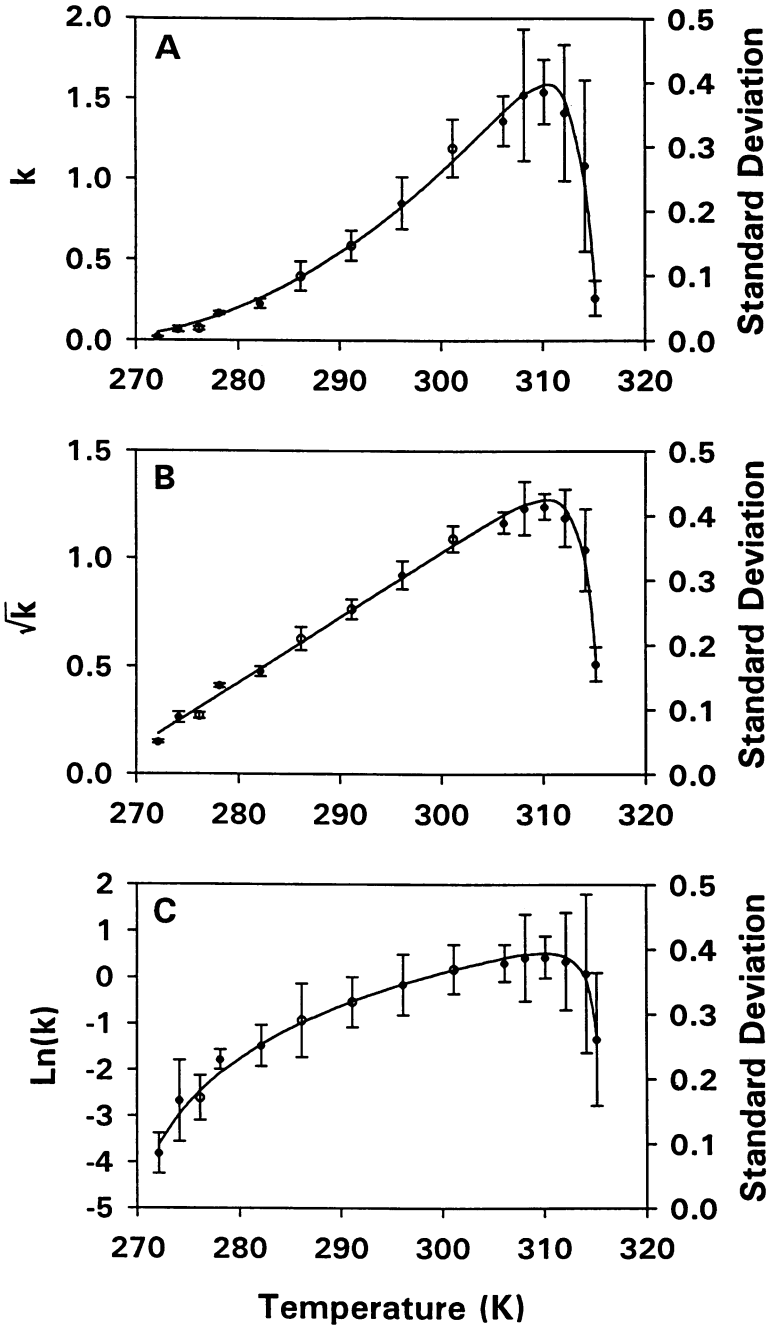

FIG. 1. Data and regression lines from square root model variates: untransformed equation 2 (A), square root transformation (equation 3) (B), and natural logarithm transformation (equation 4) (C). Error bars represent ±1 standard deviation of growth rate; ● represents a single growth rate determination; ○ represents an average of two or more independent growth rate determinations; k equals growth rate (h^{-1}).

the desired temperature. At regular intervals throughout the exponential growth phase, the concentration of a single culture tube was determined by spread plating. Cultures were diluted with a 0.1% peptone solution, and 0.1-ml samples were plated onto brain heart infusion agar in duplicate. Each tube was sampled only once. Plates were incubated at 301 K, and colonies were counted after approximately 48 h. Sampling continued until the cultures reached the stationary phase (approximately 10^9 CFU/ml).

Calculations. Growth rates were calculated from the slope (m) of the linear portion of each growth curve by the following formula:

$$k = \frac{m}{\log 2} = \frac{\Delta \log\left(\frac{\text{CFU}}{\text{ml}}\right)}{\Delta \text{hour}} \times \frac{1}{\log 2} \quad (17)$$

Variance of each slope [Var(m)] was calculated according to the following formula:

$$\mathrm{Var}(m_i) = \frac{\sum(y_i - \hat{y}_i)^2}{[\sum(x_i - \bar{x})^2](n-2)} \quad (18)$$

where y_i is the i^{th} observed concentration at time x_i [log(CFU/ml)], y_i is the i^{th} predicted concentration at time x_i [log(CFU/ml)], x_i is the i^{th} time (hours), $\bar{x}$ is the mean time (hours), and n is the number of datum points. Var(k_i) was approximated by the formula:

$$\mathrm{Var}(k_i) = \mathrm{Var}\left\{\frac{m_i}{\log 2}\right\} = \left[\frac{d}{m_i}\left\{\frac{m_i}{\log 2}\right\}\right]^2 \mathrm{Var}(m_i) = \frac{\mathrm{Var}(m_i)}{(\log 2)^2} \quad (19)$$

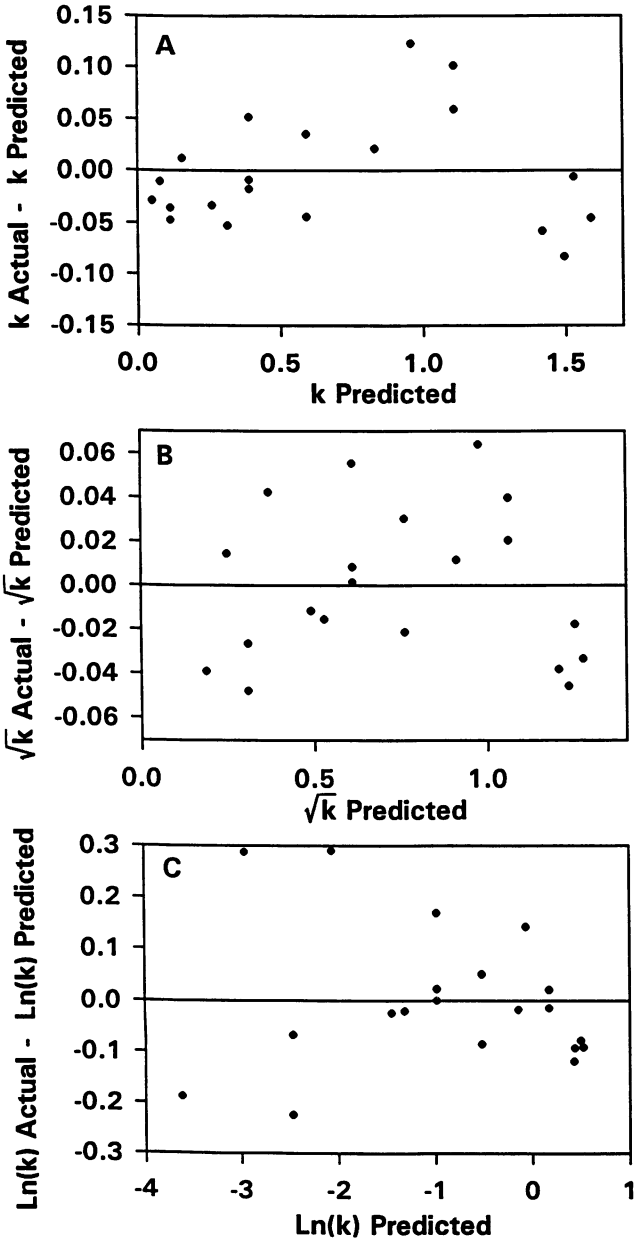

FIG. 2. Residual plots for square root model variates: untransformed equation 2 (A), square root transformation (equation 3) (B), and natural logarithm transformation (equation 4) (C). k equals growth rate (h^{-1}).

TABLE 1. Regression results[a]

Variate	Model	Equation no.	SSE
Square root	Schoolfield	6	0.0215
	Square root	3	0.0227
ln	Schoolfield	7	0.412
	Square root	4	0.356
Weighted[a]	Schoolfield	5	0.00254
	Square root	2	0.00396

[a] $[\mathrm{Var}(k)]^{-1}$ was used for calculation of weighted sum of squares due to error.

Var[ln(k)] and Var[$\sqrt{k}$] were approximated by equations 9 and 10, respectively.

Curve fitting. Parameter estimates for the Schoolfield and square root models were determined with Tablecurve 3.01 (Jandel Scientific, Corte Madera, Calif.), which uses the Levenberg-Marquardt algorithm. Starting parameter values for regression in the Schoolfield equation were calculated by methods given by Schoolfield et al. (8). Initial parameter estimates for the square root model were determined by the procedure given by Ratkowsky et al. (6).

For weighted regression, growth rate was weighted by 1/Var(k). The weights used to calculate the weighted sum of squares due to error are adjusted by the Tablecurve software so that the sum of all weights equals the number of data points (n = 21).

RESULTS AND DISCUSSION

In all cases in which the same transformation or weighting procedure was used, the square root and Schoolfield model variates produced similar regression results. The square root model variates will be used to demonstrate the statistical consequences of data transformation and weighting.

Figure 1A shows that Var(k) increases as k increases up to 310 K. The cause of this variance trend can be understood by examining the experimental method used to determine k. Cell concentration measurements were made from the lag phase to the stationary phase (approximately $10^{5.5}$ through 10^9 CFU/ml). At rapid growth rates, the organism reached the stationary phase in less time than at slow growth rates. When the growth rate is rapid, a small range of x values is used to calculate Var(m) (equation 18). This minimizes the value of the denominator, which inflates both Var(m) and Var(k) values, independent of any actual variation in experimental measurements of cell concentration.

It might be expected that when growth rate begins to decrease at temperatures above the optimum, Var(k) would decrease. However, a reduction in sampling accuracy resulted in an increase in Var(k) at temperatures above 310 K. *Y. enterocolitica* is motile at low temperatures and loses motility at temperatures above 303 K (5). At temperatures above 306 K, clumping of cells in the broth cultures made it difficult to obtain representative samples and to count colonies.

Weighted regression and data transformation were used to compensate for the heterogeneity of variance. Each model variate was evaluated primarily for its ability to correct for the deviation from the assumption of homogeneity of variance. Figure 1B and C show the effects of the square root and natural logarithm transformations on variance. The

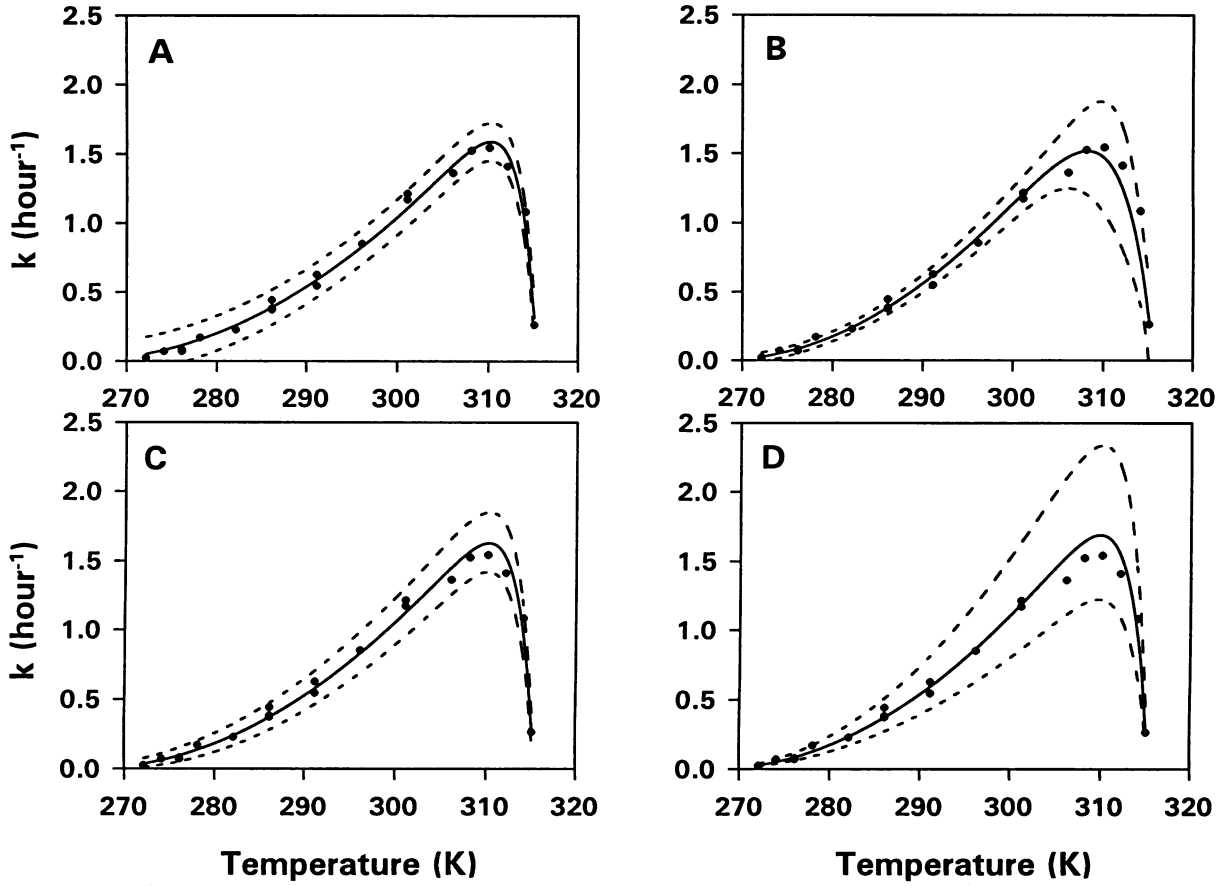

FIG. 3. Prediction intervals (- - -) for square root model variates ($\alpha = 0.05$): untransformed equation 2 (A), weighted regression (equation 2) (B), square root transformation (equation 3) (C), and natural logarithm transformation (equation 4) (D).

square root transformation diminished but did not completely remove the variance trend. The natural logarithm transformation eliminated the variance trend.

When variances are not available, residual plots are often used to assess variance homogeneity. Figure 2 shows the residual plots for the three square root model variates. The ability of residual plots to accurately reflect variance is limited by the inherent inaccuracy of the model's predictions

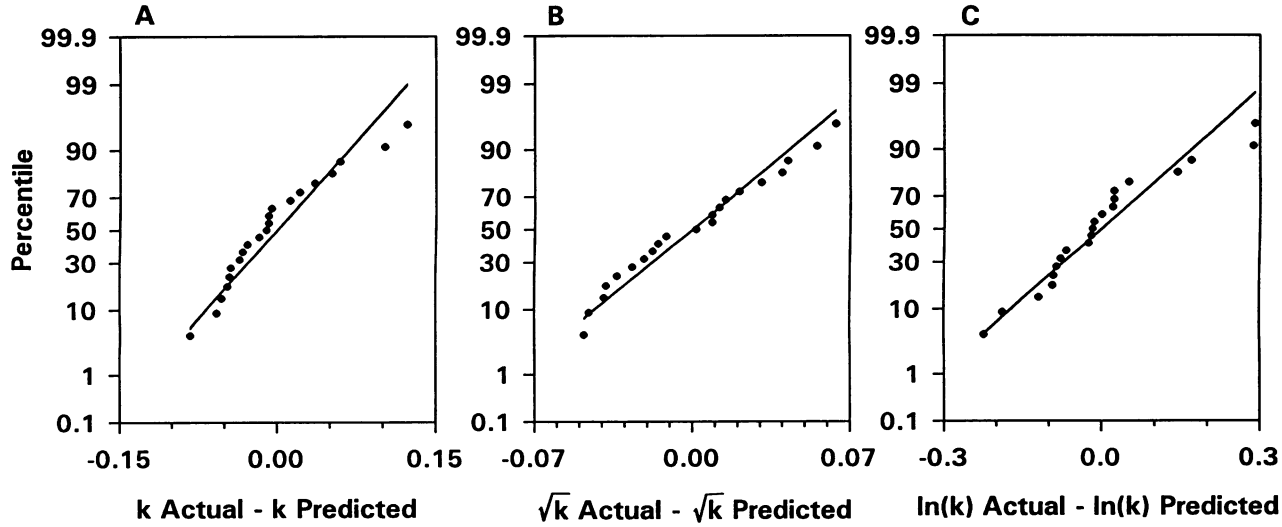

FIG. 4. Normal probability plots for square root model variates: untransformed equation 2 (A), square root transformation (equation 3) (B), and natural logarithm transformation (equation 4) (C). k equals growth rate (h^{-1}).

of the true growth rate at each experimental temperature. In Fig. 2A, the shape of the residual plot correctly reflects the variance trend. However, examination of the residual plots in Fig. 2B and C might result in the erroneous conclusion that the square root transformation results in more homogeneous variances than the natural logarithm transformation.

The regression results for all model variates examined are summarized in Table 1. Because of the difference in scaling, the SSE cannot be used to make comparisons between models with different transformations. For square root-transformed models, the SSEs for the square root and Schoolfield model variates are very similar. For natural logarithm-transformed variates, the SSE for the square root model (0.356) is lower than the SSE for the Schoolfield model (0.412). A weighted SSE is used to compare the weighted regression models. The weighted SSE for the Schoolfield model variate (0.00254) is lower than that for the square root model variate (0.00396).

The shape of the prediction intervals around the regression function is an important point to consider when evaluating models with different transformations or weighting. Prediction intervals should reflect the ability to accurately predict growth rate. Figure 3 shows the effect of weighting and data transformation on prediction intervals after data conversion back to the linear scale. An increase in the width of the prediction intervals as k increases occurs with both the square root and natural logarithm transformations and with weighted regression. This is consistent with the increase in variance as k increases.

Data transformation alters the shape of the population distribution associated with each observation. Ideally, a variance-stabilizing transformation should also result in normal population distributions. Normal probability plots, shown in Fig. 4, were used as a diagnostic tool for assessing the shape of the population distributions before and after transformation. Interpretation of these plots must be made cautiously, because of the small number of datum points. A normal probability plot will be a fairly straight line when the assumption of normality is met (1). Figure 4A indicates that the growth rate populations may be skewed right. Bacterial population distributions are frequently Poisson distributions (4). When mean values are small, Poisson distributions are skewed right. After conversion back to a linear scale, the regression lines from both the natural logarithm- and square root-transformed model variates have asymmetric prediction intervals, with the larger portion above the predicted value. This is consistent with right-skewed population distributions.

Conclusions. When the same transformation or weighting procedure was applied, the square root and Schoolfield models produced similar regression lines. Comparisons of the two models gave different results, depending on the variations of the models used. For the square root-transformed variates, the square root model had a slightly higher SSE than the Schoolfield model. However, for the natural logarithm-transformed variates, the square root model had a lower SSE than the Schoolfield model. This illustrates the importance of choosing the appropriate data transformation prior to evaluating the accuracy of different models.

The effect of data transformation on regression should be evaluated independently for each data set. If the variance of growth rate increases as the magnitude increases, either a square root or a natural logarithm transformation may correct for the variance heterogeneity. Both transformations inflate the variances of data at the low end of the scale. This effect is more extreme with the natural logarithm transformation (3). For our data set, there was considerable discrepancy between variances associated with large growth rates and those associated with small growth rates. The natural logarithm transformation eliminated this variance trend. For Smith's data, cited by Ratkowsky et al., the variance trend was less drastic and a square root transformation of growth rate was sufficient for removing the variance trend (7).

The growth rate of bacteria in food is dependent on many factors, which include storage temperature, pH, water activity, salt concentration, preservatives, oxidation/reduction potential, processing and packaging, and the presence of competitive microflora (10). The complex interaction of the effects of these factors on bacterial growth makes the development of accurate models for bacterial growth rates in foods substantially more complicated than that for a single organism in defined media. One would expect the Var(k) in food systems to be larger and more heterogeneous. Weighting or data transformation may be able to correct for the variance heterogeneity. The accuracy of the regression function prediction limits should be considered when selecting a model variate.

ACKNOWLEDGMENTS

This research was supported by state appropriations and U.S. Hatch Act funds.

REFERENCES

1. **Bates, D. M., and D. G. Watts.** 1988. Nonlinear regression analysis and its applications, p. 27, 144. John Wiley & Sons, Inc., New York.
2. **Bevington, P.** 1969. Data reduction and error analysis for the physical sciences, p. 182. McGraw-Hill Book Company, New York.
3. **Gilbert, N.** 1989. Biometrical interpretation: making sense of statistics in biology, 2nd ed., p. 64. Oxford University Press, New York.
4. **Jarvis, B.** 1989. Statistical aspects of the microbiological analysis of foods. Prog. Industrial Microbiol. **21**:36.
5. **Krieg, N. R., (ed.).** 1984. Bergey's manual of systematic bacteriology, vol. 1, p. 498. The Williams & Wilkins Co., Baltimore.
6. **Ratkowsky, D. A., R. K. Lowry, T. A. McMeekin, A. N. Stokes, and R. E. Chandler.** 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. J. Bacteriol. **154**:1222–1226.
7. **Ratkowsky, D. A., T. Ross, T. A. McMeekin, and J. Olley.** 1991. Comparison of Arrhenius-type and Bělehrádek-type models for prediction of bacterial growth in foods. J. Appl. Bacteriol. **71**:452–459.
8. **Schoolfield, R. M., P. J. H. Sharpe, and C. E. Magnuson.** 1981. Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. J. Theor. Biol. **88**:719–731.
9. **Sharpe, P. J. H., and D. W. DeMichele.** 1977. Reaction kinetics of poikilotherm development. J. Theor. Biol. **64**:649–670.
10. **Skinner, G. E.** 1991. Status summary: Moffett food safety index. Food and Drug Administration, Summit-Argo, Ill.
11. **Zwietering, M. H., J. T. De Koos, B. E. Hasenack, J. C. De Wit, and K. Van 't Riet.** 1991. Modeling of bacterial growth as a function of temperature. Appl. Environ. Microbiol. **57**:1094–1101.