e^{-K}, A FUNCTION FOR THE MODELER

Chester E. Jensen

USDA Forest Service Research Paper INT-240
INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
FOREST SERVICE, U.S. DEPARTMENT OF AGRICULTURE

USDA Forest Service
Research Paper INT-240
December 1979

e^{-K}, A FUNCTION FOR THE MODELER

Chester E. Jensen

INTERMOUNTAIN FOREST AND RANGE EXPERIMENT STATION
Forest Service
U.S. Department of Agriculture
Ogden, Utah 84401

THE AUTHOR

CHESTER E. JENSEN is Principal Statistician for the Intermountain Forest and Range Experiment Station. Previously, he held the same position at the Northeastern and Central States Forest Experiment Stations. Mr. Jensen, who holds a master of forestry degree from Michigan State University, received graduate training in statistics at Iowa State University.

RESEARCH SUMMARY

Manipulating the parameters of e^{-K}, a variant of the Normal function, generates a great variety of bell-shaped curves. These curves, or portions thereof, are useful in developing mathematical descriptors for graphed hypotheses of the relations between continous variables (regression relations). Like the Normal, e^{-K}, responds sigmoidally to departures from a pivotal value in X and to changes in the point of inflection within the X-range. Slope of the sigmoidal face at the inflection point varies additionally with change in power of the negative exponents for the e-components. All sigmoids are forced to zero at the extremities of the applicable range (the pivot point, X_p, $\pm X_p$) to enhance control of the curve system by the modeler. Within this range, e^{-K} varies from zero to one and the appropriate form of e^{-K} can be scaled to the value of the objective curve at X_p. Along with curves of the class X^n, e^{-K} is particularly useful in developing mathematical descriptors for curves with unique shapes and/or those involving complex interactions. A five-dimensional application is shown.

CONTENTS

INTRODUCTION

The December 1975 issue of Biometrics featured "A Review of Response Surface Methodology from a Biometrics Viewpoint" by R. Mead and D. J. Pike. Along with methods review, the authors examined applications in current biological literature and identified a variety of statistical improprieties. The formulation of weak hypotheses is one of these. It was noted that, when transformations are used, simple polynomials predominate and that "Polynomials seem to be used as the simplest readily available smoothing curve, without any appeal to their theoretical properties as approximations to the true response function," (p. 816-817). In the presence of adequate incentive for analysts to follow technical direction, this finding would suggest lack of emphasis on hypothesis development in statistical texts and in the curriculums of statistical schools. In any event, a brief discussion of hypothesis development is in order here as support for the presentation of e^{-K}, a family of curves designed to facilitate mathematical characterization of hypotheses that have been established graphically.

HYPOTHESIS DEVELOPMENT

A condition for valid statistical evaluation of hypotheses is that they be developed independently of the data sets used for evaluation. Analysts, then, must rely on knowledge, intuition, and conjecture to establish their concepts of the underlying forms of the relations being considered. When expressed graphically, possibly the usual case, mathematical descriptors for these relations must be established as the hypotheses to be evaluated.

There is perhaps little reason to search for exact mathematical forms to represent conceptual models regarded as weak by the analysts generating them. Descriptors based on low-degree polynomials will probably suffice.

But, as conceptual models elicit more confidence, they should be more accurately represented by mathematical hypotheses. This could necessitate a time-consuming search for appropriate *existing* functions, possibly contained in a list similar to that provided by Mead and Pike (p. 817). Direct *development* of suitable functions by the analysts, however, may prove to be a more satisfying and perhaps even a more efficient alternative. The exact nature and reliability of the graphed hypotheses are emphasized in this process and full control of the form of the descriptor is achieved by the analyst. Mathematical descriptors should meet the acceptance criteria of the analysts involved (Bartlett 1947; Draper and Hunter 1969). The parent function, e^{-K}, provides a versatile base from which to develop mathematical descriptors for forms of widely differing shapes.

Data sets reserved for evaluation of hypotheses and excluded as sources of information in the development thereof may contain information beyond that included in the original hypotheses. After evaluation, analysts are free to exhaust such data sets of new information graphically or by any means available and to incorporate these findings into advanced hypotheses. These may give rise to further research and to new data with which to evaluate advanced hypotheses.

Methods for exhausting data of information directly are shown by Jensen and Homeyer (1970, 1971) and Jensen (1973). And again e^{-K} is a transformation alternative that can be used in developing mathematical expressions for graphed hypotheses of the relations between continuous variables.

e^{-K}: ITS DERIVATION, CAPABILITIES AND LIMITATIONS

The new function, simply identified as e^{-K}, provides the analyst with a finite source of versatile transformations for use when it is inefficient to search for alternatives or when alternatives are inadequate. While e^{-K} is not a panacea for all the problems modelers face in developing functional relationships, it along with curves of the class X^n $(n \geq 1)$, does serve a broad spectrum of transformation needs with no particular limitations as to curve shapes for which it is most useful. Methods for developing mathematical descriptors using these functions have been treated by Jensen and Homeyer (1970, 1971), and Jensen (1973, 1976).

A variant of the normal, the new function is defined as:

$$e^{-K} = \frac{e^{-\left|\frac{(X/X_p) - 1}{(X_I/X_p) - 1}\right|^n} - e^{-\left|\frac{1}{(X_I/X_p) - 1}\right|^n}}{1 - e^{-\left|\frac{1}{(X_I/X_p) - 1}\right|^n}} \quad , \; 0 \leq X \leq 2X_p$$

where,

e = natural logarithm base.

X_p = pivot point in X, for e^{-K}.

I = point of sigmoidal inflection in X.

X_I = absolute departure of I from zero to X_p or, from $2X_p$ to X_p.

n = power of negative exponents for e.

The divisor X_p is retained throughout the equation for e^{-K} to preserve correspondence with the proportional X-format of the descriptor development system associated with e^{-K} in Jensen and Homeyer (1970).

The left numerator, like the Normal, generates a system of bell-shaped curves about X_p. These curves reach the maximum value of 1.0 at X_p, decline sigmoidally and symmetrically on either side of X_p with increasing absolute departure of X from X_p, and reach zero ~~at $X_p \pm X_p$~~. Sigmoids on either side of X_p are forced into different areas of two-dimensional space by shifting I, the point of inflection, and the slope of the sigmoidal face at I, changes with n.

User control of the curve system is enhanced if all curves of the set range in value from zero to one within a finite domain of X. To achieve this property, curves generated by the left numerator have been forced through zero at $X_p \pm X_p$ as follows:

Let the left numerator = e^{-T} and let $e^{-T_o} = e^{-T}$ at X = 0 or $2X_p$. Consider the residuals generated by $1-e^{-T}$. Expanded by the inverse $(1-e^{-T_o})^{-1}$ and subtracted from one, we have

$$e^{-K} = 1-(1-e^{-T_o})^{-1}(1-e^{-T})$$

which simplifies to the final form,

$$e^{-K} = (e^{-T} - e^{-T_o})(1-e^{-T_o})^{-1}$$

Then, the right numerator and the denominator serve to force curves of the left numerator through zero at $X_p \pm X_p$. This domain, along with X_p, can be altered to accommodate skewed conceptual models by adding constants (+ or -) to the X-scale. The apparent complexity of e^{-K} is much reduced by the fact than in application the right numerator and the denominator reduce to constants or approach zero and can be deleted.

Sample arrays of curves from the left half of this function (a mirror image of the right half) are shown in figure 1. Sets are shown for n = 1.5, 3.0, 5.0, and 10.0. Each set has curves progressing from $X_I/X_p = 0.1$ at the left to 0.9 at the right. It is evident that a great variety of sigmoid (and bell-shaped) curves can be generated with e^{-K}. These curves or any portions thereof provide an almost endless potential for matching graphed curves and describing them mathematically. Given that Y_p is the peak value of Y (the dependent variable) on the objective curve, the selected e^{-K} function may be scaled to that curve through multiplication by Y_p.

A descriptor (X_T), adopted as a hypothesis in its entirety, may be fitted to pertinent data by least squares in the simple model,

$Y = \beta X_T + \varepsilon$, where the ε are NID $(0, \sigma^2)$, constant variance, and $\hat{\beta} = \Sigma X_T Y/\Sigma X_T^2$

Weighted regression procedures are recommended where the variance about regression is *not* uniform over the ranges of the independent variables. In such cases, reasonable success in achieving constant variance has been obtained by solving for departures $(Y_i - \hat{Y}_i)^2$ or d_i^2 as a suitable function of the related $\hat{Y}_i$ in $d^2 = b\hat{Y}^n$. Then the weight, w, for each observation is set equal to $1/\hat{Y}^n$ and a weighted β in $Y = \beta\hat{Y}$ is estimated as $\hat{\beta} = \Sigma w\hat{Y}Y/\Sigma w\hat{Y}^2$.

More complex, generally iterative fitting procedures, such as the Newton-Raphson method, can be used to arrive at statistical estimates of internal model parameters (*see* Damaerschalk and Kozak 1977; Draper and Smith 1966).

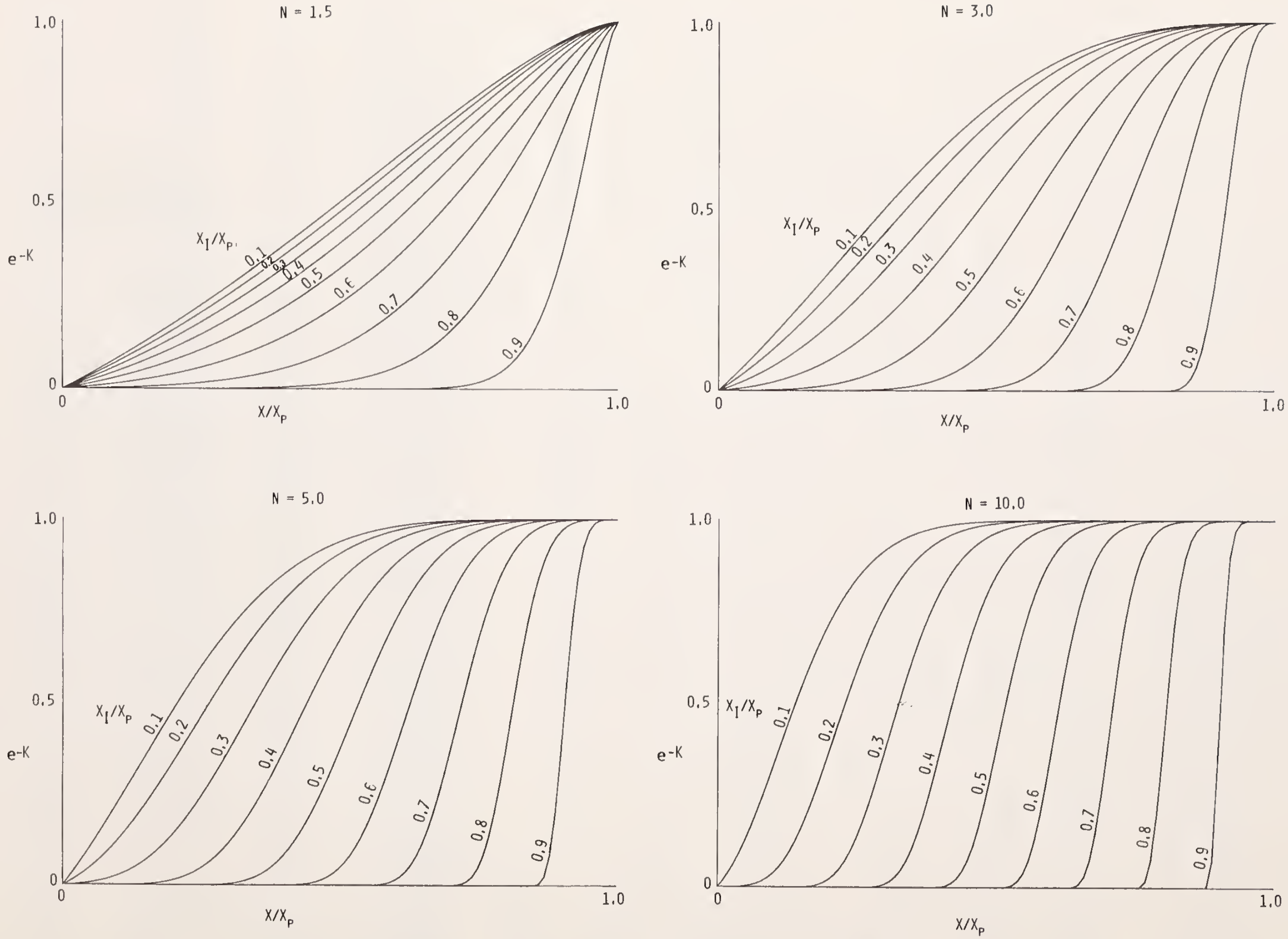

Figure 1.--Sample arrays of curves from e^{-K}.

e^{-K}: A FIVE-DIMENSIONAL APPLICATION

The flexibility of e^{-K} in representing a complex relation is evident in the five-dimensional interaction pictured in figure 2. Here an index to intensity of deer use of forest openings created by clearcutting in western Montana is expressed as a function of opening size, height of new vegetation, depth of slash in the opening, and depth of dead and down timber adjacent to the opening. (Data were provided by L. Jack Lyon, Wildlife Research Biologist, Forestry Sciences Laboratory, Intermountain Forest and Range Experiment Station, Missoula, Montana).

The data at hand, however, were initially committed to statistical evaluation of the linear effects of the above and other independent variables on the intensity index. These were virtually the simplest regression hypotheses that could have been developed. In this case, the evaluation provided only weak support for expected results and yielded little new information.

At this point, an advanced hypothesis was developed for the set of four independent variables above judged to be of high utility to the land maanger. Prior knowledge on the forms of the relations, including interactions, between these variables was summarized. Subject to these constraints and adhering to the data-fitting principles of "least deviations" (Karst 1958), the data were then graphically exhausted of associated curve form and scaling information and appeared to provide strong support for the dynamic interaction anticipated. Procedures specified by Jensen and Homeyer (1970, 1971), and Jensen (1973, 1976) were used to develop a functional form X_T for the graphed interaction. This was refitted to the data by least squares in the model $PGR = \beta(X_T) + \varepsilon$ and $\hat{\beta} = 0.9721$. The adjusted model is specified below:

$$PGR = f\ (VI, SI, SO, Acres)$$

where,

PGR = number of deer pellet groups per acre inside the opening.

VI = height of vegetation in feet, inside the opening.

SI = depth of logging slash in feet, inside the opening.

SO = depth of dead and down timber, in feet, outside the opening.

Acres = size of opening in acres.

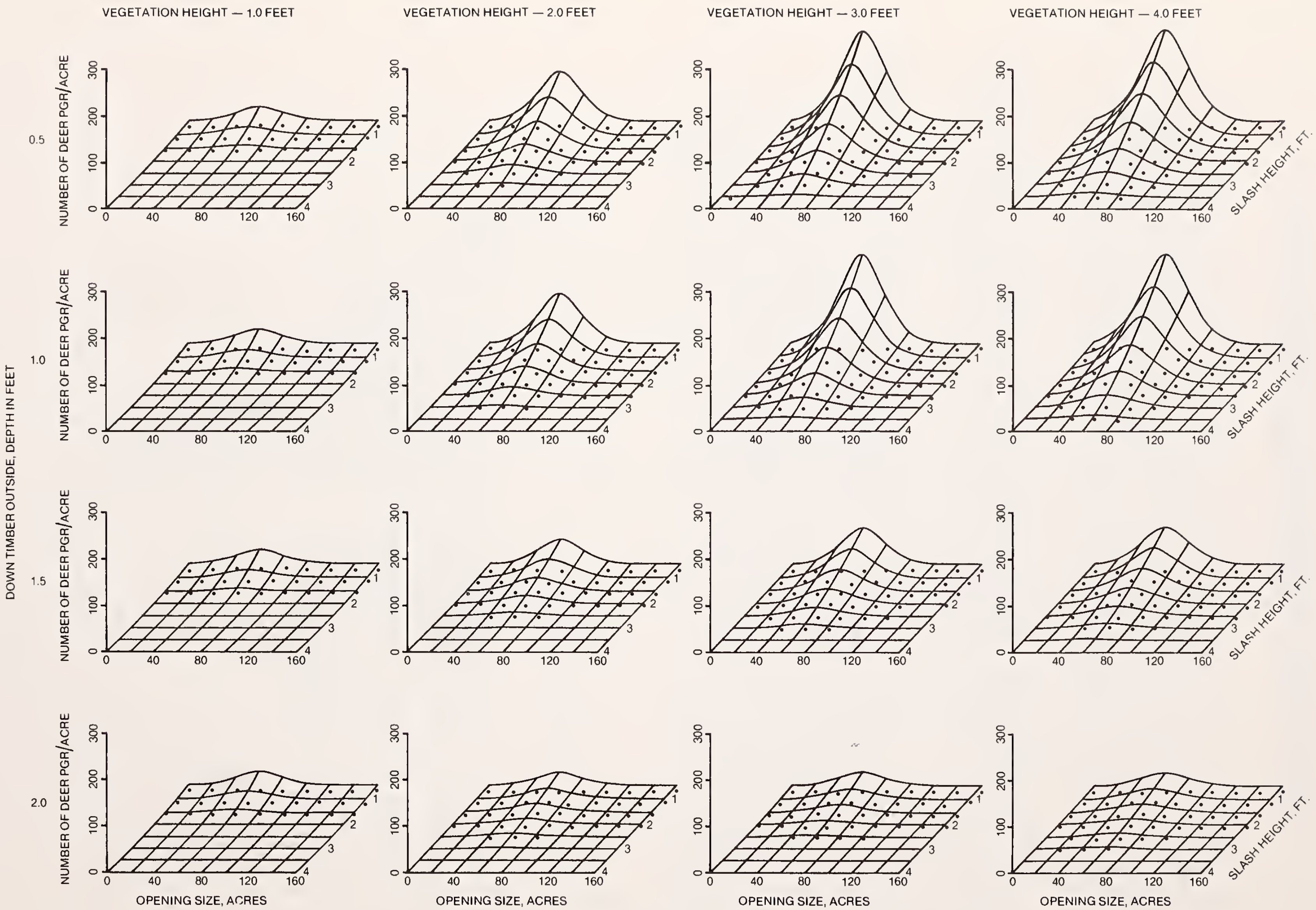

Figure 2.--Intensity of deer use of clearcut forest openings in western Montana, PGR = pellet groups per acre.

PGR = f1 (VI, SI, SO, Acres)

IF SI ≤ PO

$$PGR = \left\{ \frac{Y_p}{(PO - 0.6)^{1.55}} \left| PO - SI \right|^{1.55} \right\} (0.9721)(50)$$

where

$$Y_p = Y_pS \left\{ \frac{ e^{-\left| \frac{\frac{Acres + 200}{260} - 1}{(X_I/X_p) - 1} \right|^{1.75}} - e^{-\left\{ \frac{1}{(X_I/X_p) - 1} \right\}^{1.75}} }{ 1 - e^{-\left\{ \frac{1}{(X_I/X_p) - 1} \right\}^{1.75}} } \right\} + 0.23$$

$$Y_pS = 0.57 + 3.23 \left\{ e^{-\left| \frac{\frac{VI}{8} - 1}{0.76} \right|^{20}} \right\} \left\{ e^{-\left| \frac{\frac{3 - SO}{3} - 1}{0.492} \right|^{10}} \right\}$$

$$PO = 6.289\, e^{-\left| \frac{\frac{VI}{10} - 1}{0.999} \right|^{10.5}} - 2.289$$

$$X_I/X_p = 0.897 - 0.067\, e^{-\left| \frac{\frac{VI}{8} - 1}{0.404} \right|^{3.5}}$$

IF SI > PO

~~DI~~ PGR = 0

Limits

$0 \le SI \le 4,\quad 0 \le SO \le 3$

$0 \le VI \le 8,\quad 0 \le Acres \le 300$

$R^2_{lin} = .21,\ R^2_{x_T} \doteq .71$

Graphic development of the model described by this function apparently resulted in great sensitivity to the interaction information contained in the data. But, since the degrees of freedom thereby sacrificed were unknown, conventional statistical parameters were not estimable. Some indication as to the goodness-of-fit of the functional form to the data set from which it was largely derived is provided by the proportion of the total sum of squares explained by the model $R^2 = 0.71$ here. Contrast this with the R^2 of 0.21 achieved with a minimum-effort additive regression model wherein linear effects of the same four independent variables have been fitted to the data by least squares. It would appear, at least, that sharp focus on interaction formulation was justified.

The complexity of the foregoing function may seem unwarranted but it must be recognized that mathematical constraints are necessary to the description of relationships involving such unique forms and strong interactions as are pictured in figure 2. Familiarity with the descriptor development techniques used makes it possible to interpret the form and magnitude of parameter effects from the function itself, although the net functional effect is generally more important to the user and is much more easily understood in the computer-produced graphic display (fig. 2). This has already been determined to be an excellent medium for communicating analytical results to users, land managers in this case. Also, e^{-K} has been found quite simple to manipulate on a desk-top computer.

Although new data were not available to evaluate the advanced hypothesis statistically, the model elicited intuitive confidence to the extent that it was adopted as an interim management tool.

PUBLICATIONS CITED

Bartlett, M. S.
1947. The use of transformations. Biometrics 3:39-52.

Draper, N. R., and W. C. Hunter.
1969. Transformations: some examples revisited. Technometrics 11:23-40.

Draper, N. R., and H. Smith.
1966. Applied regression analysis. p. 277-280. Wiley and Sons, New York.

Demaerschalk, J. P., and A. Kozak.
1977. The wholebole system - a conditioned dual equation system for precise prediction of tree profiles. Can. J. For. Res. 7(3):488-497.

Jensen, C. E.
1976. Matchacurve-4: segmented mathematical descriptors for asymmetric curve forms. USDA For. Serv. Res. Pap. INT-182, 16 p. Intermt. For. and Range Exp. Stn., Ogden, Utah.

Jensen, C. E.
1973. Matchacurve-3, multiple-component and multidimensional mathematical models for natural resource models. USDA For. Serv. Res. Pap. INT-146, 42 p. Intermt. For. and Range Exp. Stn., Ogden, Utah.

Jensen, C. E., and J. W. Homeyer.
1970. Matchacurve-1 for algebraic transforms to describe sigmoid- or bell-shaped curves. USDA For. Serv., Intermt. For. and Range Exp. Stn., 22 p. Ogden, Utah.

Jensen, C. E., and J. W. Homeyer.
1971. Matchacurve-2 for algebraic transforms to describe curves of the class X^n. USDA For. Serv. Res. Pap. INT-106, 39 p. Intermt. For. and Range Exp. Stn., Ogden, Utah.

Karst, O. J.
1958. Linear curve fitting using least deviations. J. Am. Stat. Assoc. 53:118-132.

Mead, R., and D. J. Pike.
1975. A review of response surface methodology from a biometrics viewpoint. Biometrics 31:803-851.

Jensen, Chester E.
1979. e^{-K}, a function for the modeler. USDA For. Serv. Res. Pap. INT-240, 9 p. Intermt. For. and Range Exp. Stn., Ogden, Utah 84401.

The function e^{-K} is designed to generate an infinite variety of bell-shaped curves, the characteristics of which are subject to control by the analyst. These curves or portions thereof, along with curves of the class X^n, are useful in developing mathematical descriptors for graphed hypotheses of the relations between continuous variables (regression relations). They are particularly useful when the hypotheses are of unique form and/or contain complex interactions. A five-dimensional application is shown.

KEYWORDS: regression, model, transformation, bell shaped, sigmoidal, multidimensional, curvilinear interaction, response surface.

Jensen, Chester E.
1979. e^{-K}, a function for the modeler. USDA For. Serv. Res. Pap. INT-240, 9 p. Intermt. For. and Range Exp. Stn., Ogden, Utah 84401.

The function e^{-K} is designed to generate an infinite variety of bell-shaped curves, the characteristics of which are subject to control by the analyst. These curves or portions thereof, along with curves of the class X^n, are useful in developing mathematical descriptors for graphed hypotheses of the relations between continuous variables (regression relations). They are particularly useful when the hypotheses are of unique form and/or contain complex interactions. A five-dimensional application is shown.

KEYWORDS: regression, model, transformation, bell shaped, sigmoidal, multidimensional, curvilinear interaction, response surface.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979-0-677-121/73

The Intermountain Station, headquartered in Ogden, Utah, is one of eight regional experiment stations charged with providing scientific knowledge to help resource managers meet human needs and protect forest and range ecosystems.

The Intermountain Station includes the States of Montana, Idaho, Utah, Nevada, and western Wyoming. About 250 million acres, or 90 percent, of the land area in the Station territory are classified as forest and rangeland. These lands include grasslands, deserts, shrublands, alpine areas, and well-stocked forests. They supply fiber for forest industries; minerals for energy and industrial development; and water for domestic and industrial consumption. They also provide recreation opportunities for millions of visitors each year.

Field programs and research work units of the Station are maintained in:

- Boise, Idaho
- Bozeman, Montana (in cooperation with Montana State University)
- Logan, Utah (in cooperation with Utah State University)
- Missoula, Montana (in cooperation with the University of Montana)
- Moscow, Idaho (in cooperation with the University of Idaho)
- Provo, Utah (in cooperation with Brigham Young University
- Reno, Nevada (in cooperation with the University of Nevada)