

UMAP

Modules In
Undergraduate
Mathematics
and Its
Applications

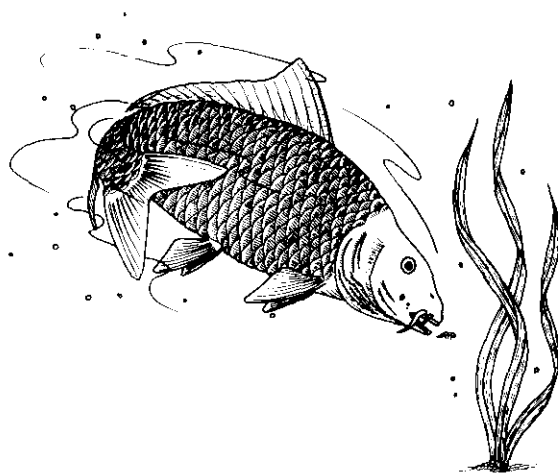
Published in
cooperation with
the Society
for Industrial
and Applied
Mathematics, the
Mathematical
Association of
America, the
National Council
of Teachers of
Mathematics,
the American
Mathematical
Association of Two-
Year Colleges, and
The Institute
of Management
Sciences.



Module 670

Least Squares, Fish Ecology, and the Chain Rule

Philip M. Tuchinsky



INTERMODULAR DESCRIPTION SHEET:	UMAP Unit 670
TITLE:	LEAST SQUARES, FISH ECOLOGY, AND THE CHAIN RULE
AUTHOR:	Philip M. Tuchinsky Research Staff Ford Motor Company Dearborn, MI 48121
MATH FIELD:	Statistics, Calculus, Differential Equations
APPLICATIONS FIELD:	Population Biology
TARGET AUDIENCE:	Students in a sophomore-level course in either multivariable calculus or differential equations.
ABSTRACT:	<i>The application:</i> Differential equation models are used to study the effect of the amount and spatial distribution of the food supply on the amount fish eat. These models are combined into a single, more encompassing model via the multivariable chain rule. The original research was by V. S. Ivlev, a Soviet biologist. His experimental data is included. The use of least-squares techniques to fit model parameters to the data is emphasized and is discussed in detail in an appendix that may be used independently as an introduction to the concepts and practice of the least-squares method.
PREREQUISITES:	Acquaintance with the multivariable chain rule, separable differential equations, calculation of extreme points of a function of two-variables, and criterion for such a point to be a minimum. No background is assumed in biology.
RELATED UNITS:	Curve Fitting via the Criterion of Least Squares (Unit 321) by John W. Alexander Jr. Fitting Equations to Data, I (Unit 433) by Thomas M. Lamm

Least Squares, Fish Ecology, and the Chain Rule

Philip M. Tuchinsky
Research Staff
Ford Motor Company
Dearborn, MI 48121

Table of Contents

1. INTRODUCTION	1
2. THE ROLE OF EXPERIMENTATION AND THE EFFECTIVENESS OF MATHEMATICAL MODELING	2
3. RATION VS. AVERAGE INTENSITY OF THE FOOD SUPPLY	5
4. EFFECTS OF PATCHINESS IN THE FOOD SUPPLY	10
5. COMBINING THE TWO MODELS	15
APPENDIX A: LEAST-SQUARES METHODS, GOOD AND BAD	
Introduction	21
A1. The Basic Least Squares Idea	21
A2. Fitting the Model $r = R(1 - e^{-at})$	25
A3. Fitting the Model $r = (R - \rho)(1 - e^{-bp}) + \rho$	28
A4. Fitting the Combined Model $r(I, p) = R(1 - e^{-aI - bp})$	29
APPENDIX B: PROPERTIES OF IVLEV'S "INDEX OF AGGREGATION" p AS A MEASURE OF PATCHINESS OF THE FOOD SUPPLY	31
REFERENCES	35
EXERCISES	36
ADDITIONAL EXERCISES TO ACCOMPANY APPENDIX A	38
SOLUTIONS TO EXERCISES	39

MODULES AND MONOGRAPHS IN UNDERGRADUATE
MATHEMATICS AND ITS APPLICATIONS PROJECT (UMAP)

The goal of UMAP was to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics and its applications to be used to supplement existing courses and from which complete courses may eventually be built.

The Project was guided by a National Advisory Board of mathematicians, scientists, and educators. UMAP was funded by a grant from the National Science Foundation and is now supported by the Consortium for Mathematics and Its Applications, (COMAP), Inc. a nonprofit corporation engages in research and development in mathematics education.

COMAP STAFF

Solomon A. Garfunkel	Executive Director, COMAP
Laurie W. Aragon	Business Development Manager
Philip A. McGaw	Production Manager
Nancy Hawley	Editorial Assistant
Katie Hellawell	Production Assistant
Annemarie S. Morgan	Administrative Assistant
Brian Sterling	Fulfillment Coordinator

UMAP ADVISORY BOARD

Steven J. Brams	New York University
Llayron Clarkson	Texas Southern University
Donald A. Larson	SUNY at Buffalo
R. Duncan Luce	Harvard University
Frederick Mosteller	Harvard University
George M. Miller	Nassau Community College
Walter Sears	University of Michigan Press
Arnold A. Strassenburg	SUNY at Stony Brook
Alfred B. Willcox	Mathematical Association of America

This module was produced in connection with a faculty workshop held at Cornell University, sponsored by the Mathematical Association of America and funded by the National Science Foundation.

1. Introduction

During the years 1939–49, the Soviet fisheries biologist Victor S. Ivlev conducted an influential study of the feeding patterns of fish, which he recorded in journal articles and a book, translated into English in 1961 [Ivlev 1961]. His work defined the discipline of trophic ecology, which he called the scientific study of the “ecological problems of the feeding of animals” [Ivlev 1961, 3].

This research involves the study of the *ration*, or amount (weight) of food eaten per animal per unit of time; thus we speak of daily, monthly, annual rations, etc. Ivlev asks, as the fundamental questions of any study in trophic ecology, how this ration depends on: (1) the amount of available food, (2) the distribution of this food in patches in the feeding space, (3) the number of animals feeding there, (4) the selecting that animals do among alternative foods. In more detail, the ration eaten by the predator is the dependent variable that is studied in terms of these independent variables:

“... the ration eaten by the predator is the dependent variable that is studied”

- (1) the intensity or concentration of available prey, i.e., the average amount of food available per unit of feeding space (area on land, volume or area in water). From experience, Ivlev expects (hypothesizes) that the ration eaten will increase as the concentration of prey increases, but will approach a saturation amount, beyond which the animal’s ration will not increase because feeding needs are satisfied. His research using this independent variable alone is the subject of Section 3 below.
- (2) the patchiness of the food supply, i.e., the extent to which the food is gathered into concentrated piles instead of being evenly spread across the feeding space. Ivlev hypothesizes that a mobile animal will travel to the parts of the feeding space where food is concentrated, will enjoy a higher availability of food there than the overall average concentration of food would suggest, and will thus feed better when food is concentrated in patches rather than being spread out evenly. Ivlev’s choice of a measure of patchiness will be explained and criticized below; it has mathematical shortcomings yet yields effective results. In terms of his chosen measurements, Ivlev shows that the beneficial effect of patchiness in the food supply is *more* important than the average intensity of the food supply in determining the ration of the fish studied by him—patchiness is not a variable to be neglected. Section 4 and Appendix B are devoted to his study of the ration in

“... patchiness is not a variable to be neglected.”

terms of patchiness alone. In Section 5, separate mathematical models using concentration or patchiness of the food as an individual variable are combined into a single more-revealing model.

- (3) the number of animals feeding in a given space. When animals compete with one another for the same food, the average ration per animal tends to decrease.
- (4) the way the animals select preferred foods from the variety available. This depends on different stages of hunger, on the size of the potential food item, and on its concealment. The results are too complex to give in a sentence or two—the interested reader is referred to [Ivlev 1961].

Ivlev's studies using intensity and patchiness of the food supply form a coherent body of knowledge and are discussed in this paper. His extensive work with the other independent variables would take us far afield and is excluded.

Two of the other papers produced at the workshop at Cornell University where this paper was written contain an interesting mathematical treatment of the question of how animals select among alternative foods [Roberts 1983; Roberts and Marcus-Roberts 1983].

2. The Role of Experimentation and the Effectiveness of Mathematical Modeling

To get answers to his questions, Ivlev conducted carefully controlled experiments instead of studying fish in their natural habitat. He defends this at some length [Ivlev 1961, 14–16], essentially saying that laws of nature are much more likely to be discovered when conditions are controlled by the scientist and one variable can be allowed to change at a time. He also points out that such careful control can only be approximated, even in an experimental setting, because of the complexity and variety of biological systems. Because truly reproducible experiments are not to be expected, he conducted each experiment twice and retained only those experiments where results of the two trials substantially agreed.

However, science depends vitally on reproducibility of results, by the original investigator and by others. Ivlev's results ought to be

reproducible, and they must be if his work is to have scientific value. His procedure at this point suffers from several defects. First, two trials is very few. Second, he does not tell us his criterion for “substantially agree.” Third, even with such a criterion, to discard data in this manner is dangerous.

We will point out other specific defects in Ivlev’s work (estimating parameters subjectively, defining too loosely his measure of aggregation, assuming tacitly no interaction between variables) and pass over others not as relevant to our purpose (giving data to an excessive number of “significant” figures, failing to specify whether data values are single observations or averages). The reader should be aware that least-squares procedures today have estimation and hypothesis testing associated to them, but Ivlev does not use currently standard methods.

Despite the improprieties and “seat-of-the-pants” flavor that his curve fitting displays, Ivlev’s modeling may be regarded as successful in pragmatic terms: his curves fit his data.

Ivlev’s results, although gotten from experiments where conditions were “unnaturally” controlled, have found practical application. In a managed fish hatchery or farm, conditions of water temperature and cleanliness, feeding population per unit feeding space, etc., can be controlled almost as well as in the laboratory. “Ivlev fish curves,” like the ones we will develop in this paper, are in fact used to predict the ration from the food supplied. Growth rates for fish are closely tied to the ration eaten (and other variables); thus, Ivlev’s work permits managers to plan the amount, type, and special distribution of the food supply so as to realize specified harvests.

“... Ivlev’s work is thought to be of excellent quality.”

Indeed, Ivlev’s work is thought by ecologists to be of excellent quality. Kenneth Watt says: “A distinguishing feature of Ivlev’s mathematical models is that they have all been validated in experimental studies with a variety of fish species: goldfish, catfish, sunfish, carp, roach, perch, bleak, bream and tench.” He calls Ivlev’s book “the most comprehensive and penetrating quantitative research on the ecology of feeding and competition . . .” [Watt 1968, 311]. A. H. Weatherly similarly mentions the “convincing” nature of Ivlev’s work [Weatherly 1972, 189].

We will particularly look at Ivlev’s use of mathematical modeling. By finding mathematical equations that fit his experimental data fairly well, Ivlev gains the two key advantages of the successful mathematical modeler:

- (1) The biological assumptions that lead Ivlev to specific types of equation-candidates are strongly supported when the candi-

dates turn out to fit the data well. (Unfortunately, Ivlev does not explain what his assumptions are!)

- (2) The equations, once obtained, provide predictions for situations where there is no experimental data.

Ivlev's mathematics will receive attention and much criticism in the pages that follow. The main points that the reader might keep in mind are these:

- (1) We will choose curves that pass near but not necessarily precisely through the data gotten from experiments. This permits us to use mathematical models that are much simpler than would otherwise be needed, and it makes sense because the experimental data inevitably contain some amount of error. (We use the method of *least squares*.)
- (2) We always want to have a sound biological rationale for the type of mathematical model chosen, and we will have such a rationale to some extent. But a model that fits the data well can be useful although poorly backed up by theory. The final test of any model, in the opinion of some investigators, is its ability to provide new insights to the subject, not the solidity of its original assumptions.
- (3) There is a natural connection between the multivariable chain rule and the methods of experimental science. Scientists often do a series of experiments using several values of one independent variable while keeping all other variables fixed. We will see that Ivlev does this with two separate variables and uses the chain rule to knit the two sets of experimental results into one model.

"The final test of any model is its ability to provide new insights to the subject,"

The classical criticism of laboratory experimentation by those who favor observations in nature is that the lab results are done in so artificial a setting that they do not transfer well to the natural setting. Ivlev gives one of the classical answers to this criticism, writing that research done in one natural setting does not necessarily transfer to another.

"... by choosing experimental methods, Ivlev gained both mathematically and practically."

The main point of this paper is that by choosing experimental methods, Ivlev gained both mathematically and practically. The chain rule usage outlined in (3) above is his major mathematical gain. In practical terms, his experimental conditions do transfer well to a technological setting such as a fish farm. They are also substantially valid for many natural settings; we outline one example in Section 5.

3. Ration vs. Average Intensity of Food Supply

Let us use this notation¹:

r = the average ration eaten, in weight (milligrams) of food per fish per unit of time. (Standard time units are chosen for each experiment.)

R = the average saturation ration, which the fish would eat if the food supply were very large in comparison to its needs; the maximal possible value of r . Units are the same as for r . R is a constant which will be found from experimentation.

I = the average intensity of the food supply, in mg of food per unit area (or volume) of feeding space. The food is spread as evenly as possible over the feeding space.

We wish to know how r depends on I . Patchiness and other variables are excluded here. From experience and common sense, we make these assumptions:

1. For $I = 0$ (no food at all), we have $r = 0$.
2. The ration eaten, $r(I)$, will increase as I increases but will not get arbitrarily large; instead, it will approach R asymptotically as $I \rightarrow \infty$. In practice, R will be approximately reached when the food supply, indicated by I , exceeds the amount that the fish can usefully consume in the time allotted.
3. The rate of growth of the ration eaten will be largest when food is most scarce (I near zero) and will decrease as I increases. Growth will be essentially zero for large I .
4. We can regard r as a smooth function.

On the basis of these ideas, Ivlev [1961, 21] selects as his mathematical model of r in terms of I the differential equation (or

¹The letters used (but not the variables) have been changed from those in [Ivlev 1961] for convenience. The reader who wishes to read further in Ivlev's 1961 book will want this conversion table:

My notation	r	R	I	p	a	A	b	ρ
Ivlev's	r	R	p	ζ	ξ	k	x	ρ

“initial value problem”)

$$\frac{dr}{dI} = a(R - r); \quad r(0) = 0 \quad (1)$$

where a is a positive constant to be found. This model does satisfy the four requirements, because

1. The correct initial value is specified.
- 2, 3. We have $dr/dI \geq 0$ because $r \leq R$ is known on physical grounds. Thus $r = r(I)$ will be an increasing function; but as r gets larger, the slope $dr/dI = a(R - r)$ decreases. Thus the curve flattens out horizontally while still rising. As $r \rightarrow R$, the slope approaches zero and the desired horizontal asymptote is achieved.
4. By picking so simple a differential equation model, we have built in the smoothness.

There are *many* other models that have these properties. For example, by the same reasoning,

$$\frac{dr}{dI} = a(R - r)f(r, I); \quad r(0) = 0$$

satisfies the four requirements for *any* choice of a smooth function f such that $f(r, I) > 0$ when $I, r \geq 0$ and decreases as I increases. Of course eq. (1) is the simplest among such models; and we do prefer a simple model (as long as it is correct). But why use a differential equation at all? We might try to write down $r = r(I)$ directly.

Ivlev does not really explain why he starts with this differential equation. He might have said that on biological grounds, r should grow at a rate proportional to its remaining “growth space before saturation” $R - r$, for that is what (1) actually says. Or he might have believed that the rate of decrease in $R - r$ should be proportional to $R - r$, i.e.,

$$\frac{d(R - r)}{dI} = -a(R - r),$$

which is equivalent to (1) (see Exercise 2). But he does not do so and these rationales are purely speculative.

Having settled on eq. (1) as a tentative statement of the manner in which r depends on I , what can we do with this model? First of

all, we can solve the differential equation to obtain the curve $r = r(I)$. Then we can decide how well data from Ivlev's experiments fit this curve, picking an appropriate ("best") value of the constant a in the process. If the fit is fairly good, we have eq. (1) as the "law of nature" satisfied by I and r , with the insight that the rate of change dr/dI is proportional to the remaining possible ration $R - r$. We also have a solution curve, which may be graphed, differentiated, rearranged algebraically (for example, we'll be able to solve for I in terms of r), etc. All of the calculus is available to help us use and understand this function $r(I)$. Most importantly, we can *interpolate*, i.e., find r for values of I between those we use in experiments.

The solution of eq. (1) is easy, because it is a separable differential equation:

$$\begin{aligned}\frac{dr}{dI} &= a(R - r), \\ \int \frac{dr}{R - r} &= \int a dI, \\ -\ln|R - r| &= aI + c, \quad (c \text{ an arbitrary constant}) \\ \ln(R - r) &= -aI - c, \quad (\text{we have used } r \leq R) \\ R - r &= e^{-aI - c} = Ce^{-aI}, \\ r &= R - Ce^{-aI}.\end{aligned}$$

Upon placing $r = 0$ when $I = 0$, we obtain $C = R$; and the solution of eq. (1) is

$$r = R(1 - e^{-aI}). \quad (2)$$

See **Figure 1** for the graph.

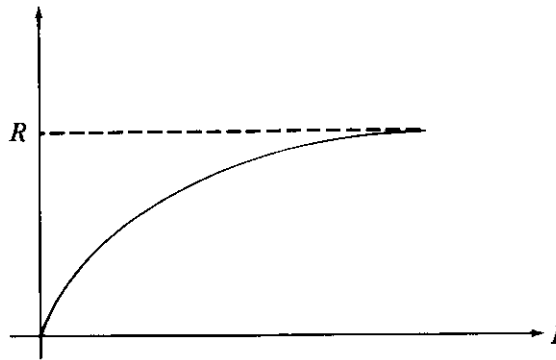


Figure 1. $r = R(1 - e^{-aI})$.

**“Ivlev connected
this mathematical
result to reality”**

Ivlev connected this mathematical result to reality by gathering (I, r) data pairs for three types of fish fed on both living and nonliving food. In each of a group of vats containing food at intensities $I = 1$ to 10 mg/cm^2 , he placed five fish that had not been fed for 18–20 hours beforehand. The weight of food eaten was recorded by “direct observation” (Ivlev does not give further details) during 1.5 to 2 hours and then confirmed by dissection of the fish’s intestines. The three series of experiments were done with

1. carp fed a nonliving diet (denatured bream² roe)
2. roach² fed on the larvae of *Chironomidae*³
3. bleak² fed on *Daphnia pulex*⁴, an actively moving prey, with the results:

Table 1.

Reprinted by permission from Ivlev 1961, p. 21.

Concentration of food objects $I, \text{mg/cm}^2$	Rations in 1st series of experiments r_1, mg	Rations in 2nd series of experiments r_2, mg	Rations in 3rd series of experiments r_3, mg
1	96.8	75.8	35.1
2	150.0	121.4	47.2
3	203.1	143.7	64.4
4	229.3	171.8	77.0
5	254.1	183.2	76.2
6	265.4	191.9	81.5
7	264.8	178.8	92.3
8	281.9	190.0	89.3
9	292.0	201.6	90.2
10	291.3	198.9	83.0

From the data of **Table 1**, Ivlev obtains values of R “by inspection,” a technique that is undesirable because it is subjective. Ivlev then uses least squares to estimate a for each Series. The resulting procedure is a strange hybrid indeed. If Ivlev is going to “guesstimate” one parameter, why doesn’t he “guesstimate” both?

²Roach, bleak, bream, and *Vimba vimba* are all European varieties of minnows (small carp), and are of commercial importance in the Baltic, Black, and Caspian Seas, thus of real concern to Ivlev.

³A tiny, non-biting harmless midge (similar in appearance to mosquitoes).

⁴A very common species of small crustacean (the 30,000 species of crustacean include lobster, crab, crayfish, etc.), in the same order (*Cladocerae*) as the water flea, found everywhere in fresh water.

Table 2.
Results of Ivlev's calculations.

Series	Subjects	Average weight of fish, mg	R	Constants	
				a	$A = a/\ln(10)$
1	Carp; nonliving food	1349	292.0	0.3887	0.1688
2	Roach; larvae	1326	198.0	0.4865	0.2113
3	Bleak; Daphnia	673	90.0	0.4414	0.1917

Much better, why doesn't he use least squares to estimate both? Appendix A discusses several least-squares approaches involving eq. (2). The Appendix shows the difficulties that Ivlev faced and indicates why he may have selected his approximation. Ivlev's results are given in Table 2. Ivlev chose to convert his formulas from use of base e to base 10 by introducing the constant A such that $e^{-a} = 10^{-A}$. (Equivalently, $A = a/\ln(10) = a \log_{10}(e)$.) His solution curves are therefore

$$r = R(1 - 10^{-At}) \quad (3)$$

and are sketched with the data points and horizontal asymptotes shown in Figure 2.

We will judge the quality of the fit (the extent to which data do obey the hypothesized mathematical model) qualitatively from this

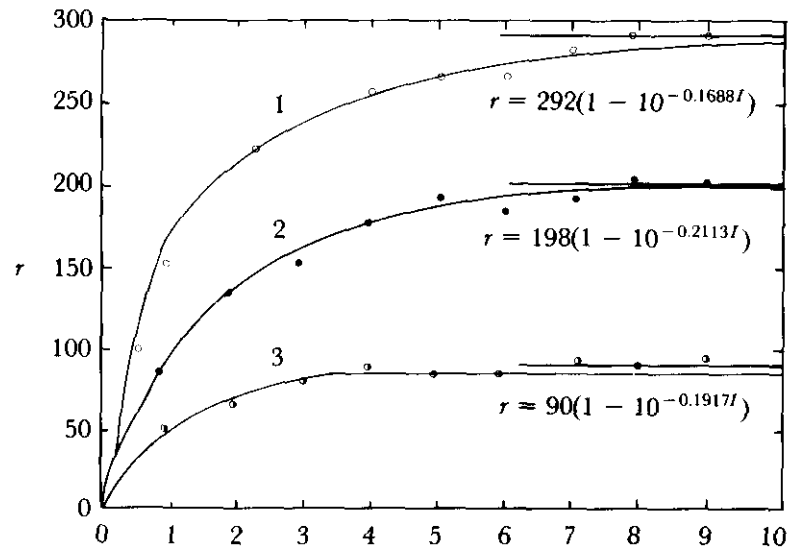


Figure 2. Reprinted by permission from Ivlev 1961, p. 22.

“These experiments support the choice of the mathematical model”

picture. See also Exercise 3. In a more advanced course that includes statistical methods associated to the least-squares process, we would perform a “regression analysis” and “analysis of variance” to obtain a mathematical measure of the goodness of fit.

These experiments support the choice of the mathematical model eq. (1), because a consistently good fit is found for three different types of fish and for living and nonliving, moving and motionless foods.

On the basis of this study, Ivlev criticizes the work of Volterra [1933; 1935], whose predation models involve unlimited growth in the ration instead of a saturation value. Volterra was a pioneer in the use of differential equations to model biological growth processes. His work has inspired many followers to propose more realistic models after his style; the interested reader might try Gause [1954] or Smith [1974]. Volterra’s simplest models have been criticized by many authors (for example, by Watt [1968, 295]) for being so unrealistically simple that they fail to match natural data.

4. Effects of Patchiness in the Food Supply

“Ivlev discovered experimentally that the more patchily a given amount of food is distributed over a fixed area, the more rapidly it is consumed.”

Ivlev discovered experimentally that the more patchily a given amount of food is distributed over a fixed area, the more rapidly it is consumed. This seems to happen because the fish travel to the patches of highly concentrated food and thus enjoy a higher local intensity of food supply at these patches than the overall average intensity of food would indicate. The effects seemed so dramatic that Ivlev set out to measure them in detail.

The main difficulty was the choice of a measurable variable that would indicate the extent of patchiness. He wanted a variable that would be zero when food was spread evenly and would increase as the same amount of food was placed in more concentrated clumps in the same space. Dependence of τ on this variable would then be studied.

“The main difficulty was the choice of a measurable variable that would indicate the extent of patchiness.”

He finally settled on what he calls the *index of aggregation*. Since a known total amount of food is to be spread unevenly over a known space, the average intensity of food supply I is known. Ivlev now split the space into n smaller regions and allocates the food supply among them with different local average intensities $I_1, I_2, I_3, \dots, I_n$. The extent of patchiness is now indicated by the differences $I_1 - I, I_2 - I, I_3 - I, \dots, I_n - I$, some of which are positive, while others are negative. These are gathered into a measure of the patchiness,

the *index of aggregation*

$$p = \sqrt{\frac{\sum_{j=1}^n (I_j - I)^2}{n}}. \quad (4)$$

Thus, the differences $I_j - I$ are squared to make them all positive contributors to the measure, and these squared differences are averaged. The square root corrects for the squaring done earlier, so that p has units of mg of food per animal per unit of time, just as I does. The reader who knows some elementary statistics should notice that p is simply the standard deviation of the sample I_1, I_2, \dots, I_n of local values of I .⁵

This measure of patchiness does have the desired properties. If food is spread evenly, $I_1 = I_2 = \dots = I_n = I$ and $p = 0$. As the food is piled into more and more locally concentrated clumps, the values I_j spread out and the squared differences $(I_j - I)^2$ increase. Then p also increases.

“However, there are serious mathematical questions about the definition of p .”

Ivlev models the dependence of r on p and publishes experimental results that confirm the general quality of the fit between data and theoretical model. However, there are serious mathematical questions about the definition of p . The main issue is that very different values of p will be calculated for the same food distribution if we choose the n subregions in different ways. Variable p depends not only on the spatial distribution of the food and on n (as we expect) but also on the shape and size of the subregions. As a mathematical quantity, it is not well-defined.

Ivlev was aware of the difficulty to some extent, for he says that the n subsections should have equal areas and should

correspond with the area covered by a single feed of the given animals. If sections are taken with measurements less than those of a single feed, the values obtained for the index of aggregation will be higher than the true biological values, whereas in the reverse instances the values of p will be correspondingly understated.

This is the only insight Ivlev offers as to what he means by “true biological value”; we must take it as synonymous with the value obtained for p by making subregions correspond to the space appropriate for a single feed.

In Appendix B, we will explore in more detail the difficulties with the “index of aggregation.” Ivlev’s point that taking more

⁵The reader may be used to formulas for the standard deviation of a sample selected randomly from a population of data, with $n - 1$ instead of n in the denominator; (4) is the standard deviation formula used by statisticians when I_1, \dots, I_n form the whole population, not just a sample. When n is large, the difference made by n vs. $n - 1$ is very minor.

subregions of smaller size tends to increase the value of p will be illustrated concretely. We will also show that the value of p depends not just on the size of the subregions but also on their shape and arrangement.

Since Ivlev does not explain how he split the feeding space into subregions in his experiments, his description of them is fatally incomplete. But he gets interesting results that are analogous to these described above in Section 3. In this section, we will present Ivlev's research without dwelling further on the qualities of p .

Ivlev sought a mathematical model for the dependence of r on p , when the average intensity of food supply I (and thus the total amount of food in the feeding space) is kept fixed but is distributed in various ways to achieve a range of values of p . He reasoned that r should equal some initial value ρ when $p = 0$ (in fact, $\rho = r(I)$, since we have studied the case of evenly spread-out food above), and that r should increase as p increases until R is approached asymptotically at saturation. Since he believed that increasing p amounts to providing the fish with a higher (local) value of I , because the fish will travel to the areas of high food intensity, he selected a near-copy of model (1) for the dependence of r on p :

$$\frac{dr}{dp} = b(R - r); \quad r(0) = \rho. \quad (5)$$

This choice is also open to criticism. A curve of the same general shape as that found for $r = r(I)$ in Section 3 is expected; however, there is no apparent rationale (to this writer and others with whom he has pursued the question) for adoption of the same model. We adopt this same model as a tentative and simple first choice; we shall be amazed at its successes.

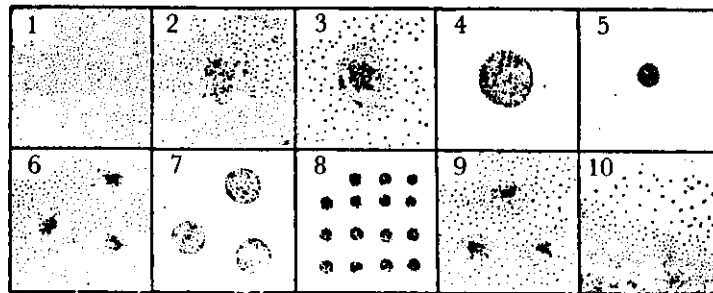
“We adopt this same model as a tentative and simple first choice; we shall be amazed at its successes.”

Notice that in (5), r is treated as a function of p alone with I held constant; we shall shortly treat dr/dp as the partial derivative $\partial r/\partial p$ that it really is.

We solve (5), regarding b , R and ρ as constants. The reader should do this in detail (Exercise 4), getting this result:

$$r = (R - \rho)(1 - e^{-bp}) + \rho. \quad (6)$$

To test this mathematical result against nature, Ivlev ran two series of experiments similar to those of Section 3. In the first series, carp were fed nonliving food (denatured roe); and in the second, carp were fed chironomid larvae. To provide a range of values of p , 10 patterns of food distribution were designed to yield p ranging roughly from 0 to 10. These patterns are exhibited and described in **Figure 3**. This was done separately for the two series, because it is difficult to obtain equal values of p for different types of food. The resulting values of r in the table of **Figure 3** are percentages of the



No.	Type of aggregation	1st series		2nd series	
		p	r	p	r
1	Absolutely even distribution of food	0	48.8	0	30.1
2	One aggregation of moderate density with indistinct outlines	1.49	71.9	1.32	65.7
3	One aggregation of increased density with indistinct outlines	4.28	91.0	4.11	93.3
4	One aggregation of increased density with sharp outlines	5.78	99.6	4.99	97.8
5	One aggregation of high density with sharp outlines	9.75	98.2	7.66	99.0
6	Three aggregations of moderate density with indistinct outlines	2.63	85.4	2.05	76.2
7	Three aggregations of increased density with sharp outlines	3.77	91.9	3.16	83.8
8	Fifteen aggregations of increased density with sharp outlines	2.54	64.2	2.20	65.8
9	Three aggregations of moderate density with indistinct outlines and of different sizes	2.60	83.7	—	—
10	Zonal distribution of the food	2.50	86.5	—	—

Reprinted by permission from Ivlev 1961, pp. 27-28.

Figure 3. The sketches show the 10 patterns of patchy food distribution used by Ivlev. These 10 patterns are described in the table, which also contains the data for the two series of experiments:

Series 1: Carp fed nonliving food. Series 2: Carp fed larvae.

maximal (saturation) ration R that arose in each series; they are not in milligrams as before. The first distribution pattern involves evenly spread food and thus yields the value p . The constant R is obtained by inspection of the rations r from each experimental series. The same least squares procedure used in Section 3 to best calculate the constant a can be used here to find b (see Appendix A, Section 3). The values obtained for p , R , and B are summarized in Table 3, and the solution curves are sketched with the actual data points in Figure 4.

Table 3.
From Ivlev 1961, p. 29.

Series	Subjects	Avg. wt. of fish grams	No. of fish in each exp.	ρ		R		b
				mg	%	mg	%	
1	Carp; nonliving food	1.02	10	104	48.8	213	100	0.465
2	Carp; larvae	4.36	5	249	30.1	826	100	0.5149

“... Ivlev can thus claim that the effects of patchiness are clearly important and can dominate the effects of the average amount of food,”

The fit is nowhere near the quality obtained in Section 3; the data points corresponding to pattern 8, shown in the dotted box on the graph in Figure 4, are particularly off-curve. However, the overall pattern is correct; and equations (5) and (6) serve as a crudely adequate model of the phenomenon. Ivlev rejects the pattern-8 data because the 15 small clumps of concentrated food involve calculation of I_1, \dots, I_n over areas that are much smaller than the carp's usual area of feeding; thus, values of p obtained were too large, and the points in the dotted box really should be shifted horizontally to the left (same r for smaller values of p) toward the curves.

The coefficients b turn out to be slightly larger than the comparable values of a gotten in Section 3. This result might be due to

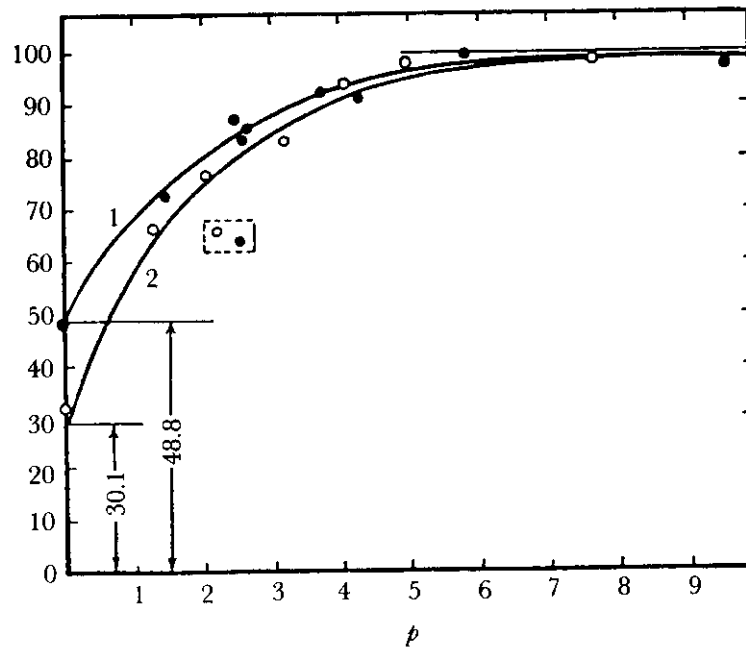


Figure 4. Curves 1 and 2 correspond to Series 1 and 2. Reprinted from Ivlev 1961, p. 28, by permission.

chance alone but the statistical tests that would tell us that cannot be performed, due to the nonstandard way in which Ivlev has developed his model. Despite this drawback, and despite our concerns about whether p is well-defined, Ivlev nonetheless claims that the effects of patchiness are clearly important and can dominate the effects of the average amount of food, I . His analysis of r in terms of p is an original and important contribution in his work; he says that “the problem of the influence exerted on intensity of feeding by the nature of the distribution of food material [thus] arises in ecology for the first time” [Ivlev 1961, 30].

He also offers an example of “nature’s knowledge” of the effect of patchiness on rations. Noting that fish spawn are the prey of many enemies, Ivlev claims that “as a rule unprotected spawn is [sic] scattered by the fish [parents] over a wide area and more or less evenly,” so that values of both I and p are low, while “spawn deposited in the form of dense conglomerations (as in the case of salmon, sheatfish, sticklebacks, etc.), is [sic] concealed or protected by the parents” [Ivlev 1961, 31]. Thus evolution has led to the qualitative use of the laws of nature uncovered here. He also offers the example of *Vimba vimba*², which distributes spawn evenly when predators are in the vicinity but in clumps when predators are absent.⁶

5. Combining the Two Models

We have studied the dependence of r on I with p held constant (in fact, with $p = 0$) and on p with I held constant. One important advantage of having these mathematical results is that they can be combined into a single model of $r(I, p)$, showing the combined effects of both variables. The experimental data by itself do not permit such a larger picture.

Since one variable was held constant while the others varied in Sections 3 and 4, we could reinterpret eqs. (1) and (5) as statements about the multivariable function $r = r(I, p)$

$$\begin{aligned}\frac{\partial r}{\partial I} &= a(R - r), \\ \frac{\partial r}{\partial p} &= b(R - r).\end{aligned}\tag{7}$$

⁶Modern biologists do not attribute intelligence and purpose to evolution, nor do they personify nature.

The initial values become

$$\begin{aligned} r(0, 0) &= 0 \\ r(I, 0) &= r(I) = R(1 - e^{-aI}) \quad \text{for any particular } I. \end{aligned} \quad (8)$$

The multivariable chain rule now says that

$$\begin{aligned} dr &= \frac{\partial r}{\partial I} dI + \frac{\partial r}{\partial p} dp \\ &= a(R - r) dI + b(R - r) dp. \end{aligned} \quad (9)$$

We are lucky enough to achieve integration easily:

$$\begin{aligned} \frac{dr}{R - r} &= a dI + b dp, \\ -\ln(R - r) &= aI + bp + c \quad (\text{for constant } c), \\ R - r &= Ce^{-(aI + bp)}, \\ r &= R - Ce^{-(aI + bp)}. \end{aligned}$$

The initial value $r(0, 0) = 0$ implies $C = R$, so we get

$$r(I, p) = R(1 - e^{-(aI + bp)}). \quad (10)$$

The first step may seem suspicious to readers who are not experienced with multivariable integration, but eq. (10) is easy to get by another method. Let's argue that through:

Since $\partial r / \partial I = a(R - r)$, integration gives

$$r = R - C(p)e^{-aI} \quad (11)$$

by the same reasoning used in Section 3; the constant of integration there has become a function of p here because of the partial derivative: each fixed value of p yields a constant of integration, which might be different for different values of p . Now differentiate eq. (11) with respect to p , holding I constant, and apply eq. (7) and then eq. (11):

$$\begin{aligned} \frac{\partial r}{\partial p} &= -\frac{dC}{dp} e^{-aI} = b(R - r) \\ &= bC(p)e^{-aI}. \end{aligned}$$

Thus the equation satisfied by $C(p)$ is

$$\begin{aligned} \frac{dC}{dp} &= -bC \\ C(p) &= ke^{-bp}, \quad \text{where } k \text{ is a constant.} \end{aligned} \quad (12)$$

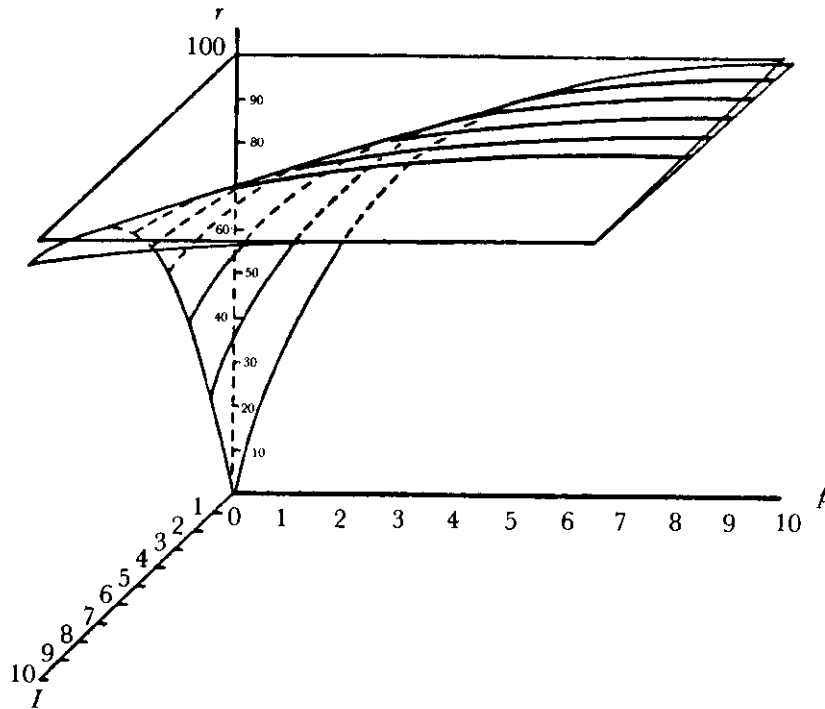


Figure 5. Reprinted by permission from Ivlev 1961, p. 34.

The notation dC/dp is correct, by the way, because C is a function of *one* variable. Putting eq. (12) into eq. (11) yields

$$r = R - ke^{-(aI+bp)}$$

and the initial value $r(0,0) = 0$ gives us back eq. (10), as promised.

A completely different combination of the two models into one, again yielding eq. (10) and using algebra without calculus, is covered in Exercise 5.

It is important to note that the derivations we have followed assume that there is no interaction effect between I and p , so that $\partial r/\partial p$ is independent of I and $\partial r/\partial I$ is independent of p . Ivlev does not mention or justify this assumption. One reason to combine the two models in the fashion he does can be to check for such interaction: If the separate fits are good but the combined fit is not, the neglected interaction may be the explanation.

Equation (10) makes sense for (I, p) in the first quadrant. As either or both of I, p move to ∞ in that quadrant, the ration values r approach R . The surface represented by (10) is pictured in Figure 5 along with the plane $r = R$ that is asymptotically approached. (Actually, all data will now be given in terms of percentages of R , so the asymptote is at height 100 above the plane. The actual function

in the figure is

$$r = 100[1 - e^{-(aI + bP)}],$$

with a and b from experimental Series 1, and is gotten by taking $r/R \times 100$ from eq. (10) to achieve percentages of R .) This surface is in fact made up of straight line segments that are all parallel to the (I, p) plane, as we learn from Exercise 6. The whole surface can thus be traced by moving a straight line in space in the right way; such a surface is called a *ruled* surface by mathematicians. A considerable theory of ruled surfaces is known.⁷

To test his mathematically constructed solution eq. (10) against nature, Ivlev began by running two series of experiments similar to those described earlier. In the first series, carp were fed denatured roe, a nonliving food. For *each* value $I = 10, 20, 30, 40, 50, 100$ mg/100 cm², Ivlev created vats with this intensity of food arranged to give a variety of values of p ranging from $p = 0$ (food spread out evenly) to a value of p between 7 and 10. For each value of I , p was taken large enough to produce near-saturation. Distribution patterns of types 2 and 3 were used (see Figure 3); these involve piling the food in a central pile that tapers off gradually and symmetrically as we move away from the center. Ten fish were used per vat and all 31 experiments (see table below) were performed twice. The data, with observed values of r expressed as a percentage of R , follow:

Table 4.
Observed values of r for Series 1: carp fed on nonliving food.
Reprinted by permission from Ivlev 1961, p. 34.

$I = 10$		20		30		40		50		100	
p	$r(I, p)$	p	r	p	r	p	r	p	r	p	r
0	22.6	0	38.5	0	55.8	0	63.9	0	67.4	0	92.3
0.90	46.2	0.59	56.2	0.94	67.0	0.95	78.0	1.41	88.5	-	-
1.82	67.6	1.48	66.0	1.36	76.9	2.09	86.3	3.34	94.3	-	-
2.79	83.8	2.35	82.2	3.24	90.0	3.21	95.4	4.65	95.7	-	-
5.21	92.7	4.35	93.6	5.33	96.6	5.0	95.8	6.66	100.0	-	-
9.02	98.3	7.09	97.8	8.96	99.9	8.42	98.7	9.32	98.9	-	-

For $I = 100$ mg/100 cm², saturation was so nearly achieved with $p = 0$ that further experiments seemed unnecessary.

In the second series of experiments, conditions were the same, except that larger carp were fed live food, chironomid larvae. This time, $I = 50, 100, 150, 200, 250,$ and 500 mg/cm² were used, each with various values of p . Comparable data were obtained, as shown in Table 5.

⁷Books on descriptive geometry contain an elementary discussion of ruled surfaces, perhaps under the headings "warped surfaces" or "classification of surfaces." The fact that a surface is ruled helps us to visualize it.

Table 5.
Series 2: carp fed larvae.

$I = 50$		100		150		200		250		500	
p	r	p	r	p	r	p	r	p	r	p	r
0	27.0	0	47.9	0	64.8	0	69.7	0	82.9	0	96.6
1.04	58.7	1.15	68.8	1.88	85.9	1.18	85.6	1.55	91.7	-	-
2.58	75.6	2.70	86.5	3.08	90.8	2.60	90.5	3.43	97.2	-	-
5.21	91.9	4.48	93.1	4.87	98.6	3.78	95.8	4.71	96.0	-	-
9.16	97.9	7.69	98.7	8.15	96.7	7.16	99.8	8.44	100.3	-	-

Table 6.
Results of Ivlev's calculations.

Series	Subjects	Average wt. of fish, g	R mg	Coefficients	
				a	b
1	Carp; nonliving food	1.98	327	0.0249	0.509
2	Carp; larvae	8.43	1205	0.00658	0.48

For each series of data, R was calculated "by inspection" and values of r were recast into percentages of R as given here. A least-squares procedure then gave "best" values of a and b for the model (10). (See Appendix A.) The values obtained by Ivlev are given in Table 6.

Figures much like those of Sections 3 and 4 are presented in Ivlev's book (pp. 35-36) and show extremely good agreement between data and model. Statisticians would use analysis of variance techniques to measure the quality of the fit, but we can get some idea of it by calculating the model predictions for a few sample points of each series⁸:

Table 7.
Experimental values vs. model predictions.

	I mg/cm ²	p	Experimental r (%)	Calculated r (%)
Series 1	20	0.59	56.2	55.0
	40	3.21	95.4	92.8
	100	0	92.3	91.7
Series 2	50	2.58	75.6	79.1
	200	0	69.7	73.2
	250	1.55	91.7	90.8

⁸For Series 1, $r = 100(1 - |\exp(-0.0249I - 0.509p)|)$.
For Series 2, $r = 100(1 - |\exp(-0.00658I - 0.48p)|)$.

The first two (I, p) pairs listed for Series 2 were visually picked as the worst fits of data points to surface. Is a mathematical model useful that gives values within about 5% of experimental data? The answer depends on the purpose of the user. The model allows calculation of any one of r, I, p when two of them are known. On a fish farm, r might be known, because a specific growth rate for the fish is wanted; and feeding methods might specify p . Then I can be predicted. When this (I, p) pair is used, actual fish rations are "experimental" values of r and should be within about 5% of the desired r value that was used to select I . The economic and biological impact of that error must still be measured.

A second complete study of the model vs. *data collected in a natural setting* was performed by Ivlev. In the Volga River delta, carp were fattened on benthos⁹ over a period of three weeks in a situation where the number of carp grew steadily but benthos distribution could be calculated per predator carp. The mathematical model produces predictions again within 5% of experimental values. Details are in Ivlev 1961, pp. 36–40.

In both the laboratory studies and the Volga River study, the typical bp products are larger than the typical aI multiplications; the extent of patchiness or aggregation of the food is a more important influence on feeding than is the overall level of food supply. The Volga study shows a situation where the available food supply decreases in total quantity but rations in fact increase, because of an increase in the patchiness variable p and the dominance of p over I . Ivlev's contention that both the spatial distribution of food and its quantity are important in an ecological study is thus supported.

The relationship between experimental methods and the multi-variable chain rule deserves a summary here. Scientists prefer to vary one variable at a time in an experiment, so that the changes that occur come from one cause, not more. (In practice this is an ideal and other variables are kept only approximately unchanged. The more creative, capable, and successful the scientific effort, the more sound the control over cause vs. effect.) If the data gotten from these one-dimensional experiments can be transformed into mathematical equations for the appropriate partial derivatives, the chain rule permits the accumulation of one-dimensional models into a single multi-dimensional one. The combined model is based on the assumption of no interaction effects among the variables; the quality of its fit will help test that assumption.

"Ivlev's contention that both the spatial distribution of food and its quantity are important in an ecological study is thus supported."

"... the chain rule permits the accumulation of one-dimensional models into a single multi-dimensional one."

⁹*Benthos* is the mixture of animals and plants living on the bottom surface of a body of water.

Appendix A: Least-Squares Methods, Good and Bad

Introduction

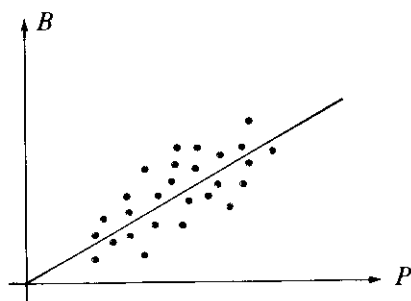
The sections that follow are intended to serve as an appendix to the UMAP Module *Least Squares, Fish Ecology, and the Chain Rule*, where the mathematical models are developed and discussed, and also to stand alone as an introduction to the least-squares method. Thus, brief summaries of the mathematical formulas from the paper are included in Sections A3, A4, and A5 for the benefit of readers not using the entire Module. Section A2 does not involve models from the paper.

A typical problem to which the least-squares method might be applied is this: We seek a mathematical formula for dependent y in terms of independent x , $y = f(x)$. Based on some theory, we believe that we know the *form* of the model. For example we might assume that the model is a straight line $y = mx + b$, or a parabola $y = Ax^2 + Bx + C$, or a sine wave $y = D \sin(Ex + F)$, or an exponential model $y = G \exp(Hx)$ where $m, b, A, B, C, D, E, F, G$, and H are *parameters* of the models. If it is the exponential model we start with, our goal is to use experimental data (x_i, y_i) , $i = 1, 2, \dots, N$, from N trials of an appropriate experiment, to choose the “best” parameters G and H for the model. We move from the form of the model to a specific formula where G, H are replaced by numbers computed from real data. To do this, we invoke a specific meaning of “best,” as will be explained below.

A1. The Basic Least-Squares Idea

Example 1: Suppose you believe or guess that B , the number of animals born per year in a colony of animals, is a constant fraction of the size P of the population of the colony: $B = mP$ for some constant m . You go out in the field and collect N data pairs (P_i, B_i) , $i = 1, 2, 3, 4, \dots, N$, by direct observation of colonies of various sizes. You know that the data (see hypothetical sketch below) do not fall precisely on any of the straight lines through the origin (remember $y = mx$?) of the (P, B) plane (if they did, any B_i/P_i would yield the needed value of m); but you also know that the data includes errors and imprecision in measurement, as all data do. You also know that $B = mP$ is probably too simple a model for such a complex process, but as a first crude model it is worthy of a look. In any case, you

have decided on $B = mP$ as your model and the task is to choose m so that $B = mP$ “best” fits the actual data (P_i, B_i) . The numbers (P_i, B_i) are known. The best line $B = mP$ will drive through the heart of the data, as in the sketch below.

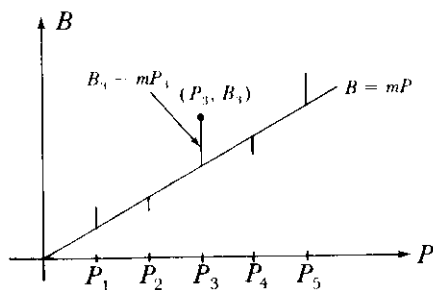


Among the many theoretically sound methods known for selecting m , the “least squares” method is by far the most used. Fix any particular constant number m in your mind. For each P_i , we have mP_i as the model’s prediction of B and B_i as actual *comparable* (i.e., related to the same P_i) values of B . Thus $mP_i - B_i$ is the error made in using the model at P_i ; we have N such errors and want to combine them into a measurement of “how good a choice m was.” The errors $mP_i - B_i$ are a mixture of positive and negative numbers; let’s square them, so they all contribute positively, and add:

$$S(m) = \sum_{i=1}^N (mP_i - B_i)^2.$$

Each choice of m yields an associated number $S(m)$; the sum S is a function of m , and a smooth one.

What does $S(m)$ tell us about “how good a choice m was”? If we choose a ridiculous value of m , the differences $mP_i - B_i$ will be fairly large in absolute value and $S(m)$ will be large. A better choice of m will lead to smaller $S(m)$ and the choice of m that makes $S(m)$ a *minimum* will yield a line $B = mP$ that drives nicely through the data.



The choice of m that makes $S(m)$ an absolute minimum for all real m is defined to be the *best value of m according to the method of least squares*. In practice, we regard m as a variable, and we seek to find the specific value $m = m_0$ that makes the function $S: m \rightarrow S(m)$ a minimum. Let's do that here by setting dS/dm to zero to seek a local minimum; in the calculation below, remember that the P_i and B_i are known numbers and only m is a variable. We differentiate term by term by the chain rule:

$$\frac{dS}{dm} = \sum_{i=1}^N 2(mP_i - B_i) \cdot P_i = 2m \sum P_i^2 - 2 \sum B_i P_i;$$

then

$$\frac{dS}{dm} = 0 \text{ gives } m = m_0 = \frac{\sum B_i P_i}{\sum P_i^2},$$

where all the sums run from $i = 1$ to N . That this is a local minimum is confirmed by

$$\frac{d^2S}{dm^2}(m_0) = 2 \sum P_i^2 > 0.$$

In Exercise A1, you are asked to show that it is also an absolute minimum for all real m .

Summary: To fit $B = mP$ to data pairs (P_i, B_i) , we set up the sum of squares errors $S(m)$ in terms of m and used differential calculus to find the $m = m_0$ that makes $S(m)$ a minimum. (Hence the name “least squares.”) This m_0 is the “best” value of m to use in the model, according to this method.

Example 2: This time we decide to fit a general straight line $y = a + bx^*$ to known data (x_i, y_i) , $i = 1, 2, 3, \dots, N$.^{A1} For each x_i , we have $a + bx_i$ and y_i as comparable values; the errors made in replacing the data by the model are $a + bx_i - y_i$ for $i = 1, 2, \dots, N$. Thus we wish to pick $a = a_0$ and $b = b_0$ such that the sum of squared errors

$$S(a, b) = \sum_{i=1}^N (a + bx_i - y_i)^2$$

is a minimum.

This function of two variables can be minimized by setting $\partial S/\partial a = 0$ and $\partial S/\partial b = 0$ as simultaneous equations. [In detail:

^{A1}A simple application: A taxi company wants a formula for its costs per day y in operating x cabs. There is a basic cost a of being in business (someone to answer the telephone, office and garage space, etc.) and there are costs b per taxicab, thus bx for x cabs. Then $y = a + bx$ is a first (very crude) model.

any local minima, maxima, or saddle points will be among the points (a, b) that make these partial derivatives simultaneously vanish. We will still need to establish that we have a local minimum and an absolute minimum.] As before, the chain rule gives:

$$\frac{\partial S}{\partial a} = \sum_{i=1}^N 2(a + bx_i - y_i)(1) = 0$$

and

$$\frac{\partial S}{\partial b} = \sum 2(a + bx_i - y_i)(x_i) = 0,$$

which we easily rearrange as

$$Na + (\sum x_i)b = \sum y_i, \tag{A1}$$

and

$$(\sum x_i)a + (\sum (x_i^2))b = \sum x_i y_i$$

after we recall that $\sum a = Na$. We thus have two simultaneous linear equations to solve for the needed a and b , most fortunately. We solve and get

$$a_0 = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum (x_i y_i))}{N\sum (x_i^2) - (\sum x_i)^2} \tag{A2}$$

$$b_0 = \frac{N \cdot \sum (x_i y_i) - (\sum x_i)(\sum y_i)}{N\sum (x_i^2) - (\sum x_i)^2}$$

as the “best” values of a and b , based on the data, for the model. (Most authors call (A1) the *normal equations* for this least squares derivation, but some refer to (A2) as the normal equations or normal solutions.) By doing Exercise A2, you will show that this is a local minimum. (It is tempting to think that for a smooth surface $S = S(a, b)$ having no other local minima, maxima, or saddle points, that (a_0, b_0) being a local minimum is enough to make (a_0, b_0) the absolute minimum as well. In fact, (a_0, b_0) is the absolute minimum, but further reasoning is required, as the tempting generalization is false [Smith et. al. 1985].)

We have again set up a sum of squared errors and used differential calculus to minimize it to find the “best” parameter values.

Exercises A3 and A4 extend the usefulness of Example 2 so that the models $y = a \cdot e^{bx}$ and $y = ax^b$ can be fit.

A2. Fitting the Model $r = R(1 - e^{-aI})$

In his book, *Experimental Ecology of the Feeding of Fishes*, V. S. Ivlev develops this model as part of a study of how much food fish eat in terms of the available food supply. Specifically, the variables are

I = average concentration of food supply (prey) in weight of food per unit area of feeding space. The food is spread evenly over the feeding space.

r = the average ration eaten in weight (mg) of food per fish per unit of time.

R = the saturation value of r ; the average maximal ration a fish would eat if the food supply were very much larger than its needs.

“We will present here four separate approaches to selecting the ‘best’ values of R and a .”

In Section 3 of my paper, *Least Squares, Fish Ecology, and the Chain Rule*, Ivlev’s research in relation to this model is discussed, including the experiments that provided data pairs (I_j, r_j) , $j = 1, 2, \dots, N$. Ivlev wanted to select the “best” values of R and a to fit $r = R(1 - e^{-aI})$ to this data. We will present here *four* separate approaches to selecting the “best” values of R and a . The first three all fail because they lead to equations for R and/or a that are impossible to solve algebraically and are ugly to solve approximately by computer. (The fact that we had no worse than simple linear equations to solve in Examples 1 and 2 above is a blessing having to do with the linearity of the models $y = mx$ and $y = a + bx$ and does not carry over to most other situations.) The fourth try will succeed and is the method Ivlev used. The reader who simply wants the successful method may skip the blind alleys that follow; however, there’s a lot to be learned about what not to do in mathematics and why, so you are invited to read straight through.

Failure #1: We do the obvious; regarding R and a both as variables, we set up

$$S_1(R, a) = \sum_{j=1}^N [R(1 - e^{-aI_j}) - r_j]^2$$

and attempt to solve simultaneously $\partial S_1 / \partial R = 0$ and $\partial S_1 / \partial a = 0$. The equations are

$$\frac{\partial S_1}{\partial R} = \sum_{j=1}^N 2[R(1 - e^{-aI_j}) - r_j][1 - e^{-aI_j}] = 0$$

and

$$\frac{\partial S_1}{\partial a} = \sum 2 \left[R(1 - e^{-aI_j}) - r_j \right] \left[RI_j e^{-aI_j} \right] = 0,$$

and rearrangement gives

$$R \sum (1 - e^{-aI_j})^2 = \sum r_j (1 - e^{-aI_j})$$

and

$$R \sum I_j (1 - e^{-aI_j}) e^{-aI_j} = \sum r_j e^{-aI_j} I_j$$

to be solved simultaneously. It is easy to eliminate R between these but the resulting equation for a in terms of the (I_j, r_j) data is horrendous. Time to quit!!^{A2}

Failure #2: Notice that a 's location as an exponent is the main problem in Failure #1. Let's bring a out in the open by using logarithms and try again: the model can be rearranged as

$$\begin{aligned} r &= R(1 - e^{-aI}), \\ \frac{r}{R} &= 1 - e^{-aI}, \\ e^{-aI} &= 1 - \frac{r}{R}, \\ -aI &= \ln\left(1 - \frac{r}{R}\right). \end{aligned}$$

The model now predicts $\ln(1 - r_j/R)$ from I . The error of the fit for (I_j, r_j) is $\ln(1 - r_j/R) - (-aI_j)$ and leads us to attempt to minimize

$$S_2(R, a) = \sum_{j=1}^N \left[\ln\left(1 - \frac{r_j}{R}\right) + aI_j \right]^2,$$

by setting two first-order partial derivatives for zero. We get

$$\frac{\partial S_2}{\partial R} = \sum 2 \left[\ln\left(1 - \frac{r_j}{R}\right) + aI_j \right] \left(\frac{r_j R^{-2}}{1 - \frac{r_j}{R}} \right) = 0$$

^{A2}Ugly, horrendously nonlinear equations arise very frequently in mathematics and there are many ways to calculate approximate solutions to them, usually by computer. The interested reader should consult texts on numerical analysis. We quit without trying such methods here because a practical way will be found that yields adequate solutions more simply.

and

$$\frac{\partial S_2}{\partial a} = \sum 2 \left[\ln \left(1 - \frac{r_j}{R} \right) + a I_j \right] I_j = 0.$$

Both equations can easily be solved for a , which can thus be eliminated between them. The resulting equation for R , however, is a nightmare of logarithms and reciprocals. Things are not working out here!

Failure #3: It looks as though least-squares fitting of both R and a is not practical (except by more advanced approximation schemes, as mentioned earlier). Fortunately, we can experimentally observe R as the saturation value that the r_j approach as I_j increases. Can we select the “best” value of a given this value of R ? Maybe so: we go back to the original model $r = R(1 - e^{-at})$ and notice that $R(1 - e^{-at})$ and r_j are comparable results from the model and data. Then we seek to minimize

$$S_3(a) = \sum_{j=1}^N \left[R(1 - e^{-aI_j}) - r_j \right]^2,$$

which looks a lot like $S_1(R, a)$ from above. However, we are now thinking of R as known, so we have a function of a alone. We set

$$\frac{dS_3}{da} = \sum 2 \left[R(1 - e^{-aI_j}) - r_j \right] R I_j e^{-aI_j} = 0,$$

so

$$R \sum I_j (1 - e^{-aI_j}) e^{-aI_j} = \sum r_j e^{-aI_j} I_j.$$

This is still an unpleasant nonlinear equation for a , so I’m going to try:

Successful calculation of a : Let’s start with R known and bring a out in the open as in Failure #2. The model is again rewritten as

$$-aI = \ln \left(1 - \frac{r}{R} \right),$$

so that this model and data (I_j, r_j) lead to comparable expressions $-aI_j$ and $\ln(1 - r_j/R)$. Thus we try to minimize

$$S_4(a) = \sum_{j=1}^N \left[\ln \left(1 - \frac{r_j}{R} \right) + a I_j \right]^2$$

by setting

$$\frac{dS_4}{da} = \sum 2 \left[\ln \left(1 - \frac{r_j}{R} \right) + a I_j \right] (I_j) = 0,$$

so

$$a \sum I_j^2 = - \sum I_j \ln \left(1 - \frac{r_j}{R} \right)$$

and

$$a = \frac{- \sum I_j \ln \left(1 - \frac{r_j}{R} \right)}{\sum (I_j^2)}.$$

With R and the (I_j, r_j) known, this a can be calculated and we succeed in selecting a “best” value of a for the model. Claim: $a > 0$ because each $I_j \ln(1 - r_j/R)$ is negative. Do you see why?

A3. Fitting the Model $r = (R - \rho)(1 - e^{-bp}) + \rho$

Ivlev goes on to analyze the influence of patchiness in the food supply on the ration eaten by his predator fish. He assumes that the fish eat a ration ρ when a given amount of food is spread evenly over the feeding space, and eat better as the same food is placed more and more in isolated clumps or patches. This is true because the fish travel to the patches of food and enjoy a larger supply per unit feeding space there than would otherwise be the case, so they eat more. The ration eaten is still r , and R is the saturation ration. The “index of aggregation” p is developed by Ivlev as a measure of the patchiness. In Section 4 of the Module, the equation in the title above is developed (and heavily criticized) as a model for r in terms of p . Ivlev’s experimental data is also presented in Section 4, and in Chapter 3 of Ivlev’s book (pp. 24–31). The experiment yields (p_j, r_j) data pairs along with known values of R and ρ . The task is to select the best value of b . We use some hard-won wisdom from Section A2 just above and rearrange the model to bring b down front at the expense of hiding R , ρ , and r behind a natural logarithm:

$$r = (R - \rho)(1 - e^{-bp}) + \rho,$$

$$\frac{r - \rho}{R - \rho} = 1 - e^{-bp},$$

$$e^{-bp} = 1 - \frac{r - \rho}{R - \rho},$$

$$-bp = \ln \left(1 - \frac{r - \rho}{R - \rho} \right).$$

This suggests that

$$S(b) = \sum_{j=1}^N \left[\ln \left(1 - \frac{r_j - \rho}{R - \rho} \right) + bp_j \right]^2$$

should be minimized. We set

$$\frac{dS(b)}{db} = \sum 2 \left[\ln \left(1 - \frac{r_j - \rho}{R - \rho} \right) + b p_j \right] p_j = 0,$$

so

$$b \sum p_j^2 = - \sum p_j \ln \left(1 - \frac{r_j - \rho}{R - \rho} \right)$$

and

$$b = - \frac{\sum p_j \ln \left(1 - \frac{r_j - \rho}{R - \rho} \right)}{\sum p_j^2}.$$

This value of b can be calculated from the available (p_j, r_j) , R and ρ . We do have a local minimum at least: you can easily show that $d^2S/db^2 > 0$.

It is worth mentioning that, if b were available from an experiment, both R and ρ could be easily chosen by the least squares method. We would seek to minimize

$$S(R, \rho) = \sum \left[(R - \rho)(1 - e^{-b p_i}) + \rho - r_j \right]^2$$

and find that two simultaneous linear equations for R and ρ emerge when we set $\partial S/\partial R = \partial S/\partial \rho = 0$. That's no help to Ivlev, however. The secret is that we have a *linear* expression in R and ρ being squared in each term above. The squaring and differentiation combine to give linear equations for R, ρ .

A4. Fitting the Combined Model

$$r(I, p) = R(1 - e^{-aI - bp})$$

In Section 5 of my paper, the models developed for r in terms of I and p separately are combined into the model in the title here. The multivariable chain rule is the mathematical mechanism used. Experiments to validate the combined model are presented there as well. Ivlev obtained (I_j, p_j, r_j) data triples. An obvious first effort to find parameters R, a, b by the least squares method would be to minimize:

$$S_1(R, a, b) = \sum \left[R(1 - e^{-aI_j - b p_j}) - r_j \right]^2$$

by setting $\partial S_1/\partial R$, $\partial S_1/\partial a$, and $\partial S_1/\partial b$ to vanish simultaneously. As you may expect from efforts in Section A3 with the special case $b = 0$, this does not succeed. The three simultaneous equations are too complicated to solve.

We opt again to obtain R from experimental observation and seek a and b by least squares. It makes sense to reverse the model

$$\begin{aligned} r &= R(1 - e^{-aI - bp}), \\ e^{-(aI + bp)} &= 1 - \frac{r}{R}, \\ -(aI + bp) &= \ln\left(1 - \frac{r}{R}\right), \end{aligned}$$

because

$$S_2(a, b) = \sum_{j=1}^N \left[\ln\left(1 - \frac{r_j}{R}\right) + aI_j + bp_j \right]^2$$

looks as if it will yield workable partial derivatives. We get

$$\frac{\partial S_2}{\partial a} = \sum 2 \left[\ln\left(1 - \frac{r_j}{R}\right) + aI_j + bp_j \right] I_j$$

and

$$\frac{\partial S_2}{\partial b} = \sum 2 \left[\ln\left(1 - \frac{r_j}{R}\right) + aI_j + bp_j \right] p_j,$$

and setting both to zero leads us to

$$a \sum I_j^2 + b \sum I_j p_j = - \sum I_j \ln\left(1 - \frac{r_j}{R}\right)$$

and

$$a \sum I_j p_j + b \sum p_j^2 = - \sum p_j \ln\left(1 - \frac{r_j}{R}\right),$$

which may indeed be solved simultaneously. The solution is

$$\begin{aligned} a &= \frac{(\sum I_j p_j) \left(\sum p_j \ln\left(1 - \frac{r_j}{R}\right) \right) - (\sum p_j^2) \left(\sum I_j \ln\left(1 - \frac{r_j}{R}\right) \right)}{(\sum I_j^2) (\sum p_j^2) - (\sum I_j p_j)^2}, \\ b &= \frac{(\sum I_j p_j) \left(\sum I_j \ln\left(1 - \frac{r_j}{R}\right) \right) - (\sum I_j^2) \left(\sum p_j \ln\left(1 - \frac{r_j}{R}\right) \right)}{(\sum I_j^2) (\sum p_j^2) - (\sum I_j p_j)^2}. \end{aligned}$$

What criterion would you use to establish that this choice of a, b is a local minimum, not a saddle point or maximum?

Appendix B: Properties of Ivlev's "Index of Aggregation" p as a Measure of Patchiness of the Food Supply

As we examined Ivlev's work on the influence of the spatial distribution of the food upon the amount eaten in Section 4, we found two key criticisms of Ivlev's choice of p as his measure of patchiness. Recall that, to calculate p , we split the feeding space into n non-overlapping subregions of equal area and find the average intensities I_1, \dots, I_n of the food in them. Let I be the overall average intensity for the whole feeding space (or, equivalently, I is the average of I_1, I_2, \dots, I_n). Then

$$p = \sqrt{\frac{\sum_{j=1}^n (I_j - I)^2}{n}}$$

The first criticism is that the value of p depends on the number of subregions chosen (or, equivalently, on their size, since we regard the total feeding space and the food distribution as fixed). Ivlev understood this and established the "true biological value" of p as the value obtained when the subregions have areas that "correspond with the area covered by a single feed of the given animals" [Ivlev 1961, 30]. We will see in Example 1 below that p does indeed change as Ivlev predicted as we vary the *size* of the subareas.

In Example 2 we will see that p also depends on the *shape* of the subregions even when they are of the same size. It also depends on the *arrangement* of the subregions even when they have the same size and shape. Ivlev makes no mention of this and does not give the shape of the subregions used in his experiments.

Example 1: Let's assume that the overall food intensity is I and that the feeding space consists of just two equal areas that each "correspond to a single feed":

$I_1 = I + x$	$I_2 = I - x$
---------------	---------------

The two local food intensities must average out to I , so they will be $I \pm x$ for some $x > 0$, as shown. For this situation, the differences from I are $I_1 - I = x$ and $I_2 - I = -x$; thus $p = \{[x^2 + (-x)^2]/2\}^{1/2} = x$. This is Ivlev's "true biological value" of p for this food distribution.

Next we leave the food distributed in exactly the same way but we create four areas by splitting each of the earlier two in equal halves. These four subareas are now smaller than the natural space of one feed, and Ivlev claims that the value obtained for p will be larger than the "true biological value" $p = x$ from above. The picture is:

$I_1 = I + x + y$	$I_3 = I - x + z$
$I_2 = I + x - y$	$I_4 = I - x - z$

The two left-hand areas must yield local intensities I_1 and I_2 that average out to the $I + x$ used before for the same space; thus, these intensities are $I + x \pm y$ for some $y > 0$. Similarly, in the two right-hand areas, I_3 and I_4 must average to $I - x$ and will thus be $I - x \pm z$ for some $z > 0$. The four differences from I are $x + y$, $x - y$, $-x + z$ and $-x - z$ and

$$\begin{aligned}
 p &= \sqrt{\frac{(x+y)^2 + (x-y)^2 + (-x+z)^2 + (-x-z)^2}{4}} \\
 &= \sqrt{\frac{x^2 + 2xy + y^2 + x^2 - 2xy + y^2 + x^2 - 2xz + z^2 + x^2 + 2xz + z^2}{4}} \\
 &= \sqrt{\frac{4x^2 + 2y^2 + 2z^2}{4}} \\
 &= \sqrt{x^2 + \frac{y^2 + z^2}{2}} > x.
 \end{aligned}$$

"The splitting into more local subareas does cause p to increase, as claimed."

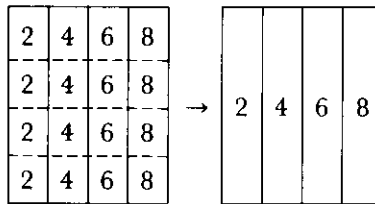
The quantity under the radical sign is greater than x^2 ; thus, $p > x$ does follow. The splitting into more local subareas does cause p to increase, as claimed.

The argument goes equally well the other way: if the four regions happen to be the natural size for one feed, then the "true biological value" of p is $[x^2 + (y^2 + z^2)/2]^{1/2}$. If we use the two-area picture, thus picking areas that are too large, we get $p = x$, a value that is unbiologically small, just as Ivlev claimed. This completes Example 1.

Example 2: A numerical example will let us see the effects of using subregions of the same size but different shape. Let the total feeding space be a square, 4 meters on a side, with an average food intensity of 5 mg per square meter. Thus there is $5(16) = 80$ mg of food in all, which we arrange in this way in subregions of one square meter each:

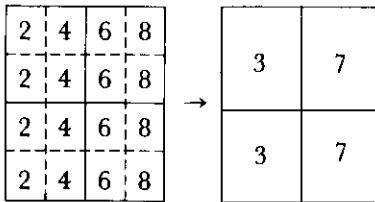
2	4	6	8
2	4	6	8
2	4	6	8
2	4	6	8

Let's assume that the natural area for a single feed is 4 square meters, so that the feeding space is to be split into four subregions. Do it this way first:



The right-hand picture shows the subregional average intensities I_1, I_2, I_3, I_4 , which are 2, 4, 6, 8, respectively. (For example, the left-most column of area 4 square meters contains 8 mg of food, or $I_1 = 8/4 = 2$ mg/m².) Recall that the overall intensity is $I = 5$; thus, differences $I_j - I$ of $-3, -1, 1, 3$ lead to $p = p_1 = \sqrt{(9 + 1 + 1 + 9)/4} = \sqrt{5}$.

Next, we choose four equal subregions of the same size but different shape:



The local intensities are 3, 3, 7, and 7 as shown. For example, the upper right-hand subregion contains $6 + 6 + 8 + 8 = 28$ mg of food, or 7 per square meter. Since I is still 5, the $I_j - I$ are $-2, -2, +2, +2$, and $p = p_2 = \sqrt{(4 + 4 + 4 + 4)/4} = \sqrt{4} = 2$.

The point is made: p depends on size *and shape or arrangement* of the subregions. But it is interesting to do one more example:

2	4	6	8	→	5
2	4	6	8		5
2	4	6	8		5
2	4	6	8		5

Each subregion contains 20 mg of food per 4 square meters of area, so the local intensities are 5, 5, 5, and 5. Since $I = 5$, the gaps $I_j - I$ are all zero and $p = p_3 = 0$.

If food were spread out evenly, *any* choice of subregions would lead to $p = 0$. (Why?) When there is in fact substantial patchiness in the food distribution, *some* choices of subregions can still yield $p = 0$ and thus totally hide the factual patchiness!

Almost any uneven arrangement of the food leads to different p as we change the size and/or shape of the subregions. For example, the two splittings below of 60 mg of food spread over 12 square meters of space involve subregions of the same shapes but different arrangement. The values of p are different, as the reader should verify.

2	4	6	8
2	4	6	8
5	5	5	5

$$p = \sqrt{8/3}$$

2	4	6	8
2	4	6	8
5	5	5	5

$$p = \sqrt{2/3}$$

The use of p is thus quite ticklish! We have to wonder how Ivlev in fact chose his subregions and whether his choice influenced the mathematical models that emerged. Does the rather good fit of Ivlev's (p_j, r_j) data (see Section 4) result from the model he chose or from luck and/or genius in selecting subregions? We cannot tell.

Yet Ivlev does establish the importance of studying patchiness. In a laboratory or fish-farm setting, he has shown that high patchiness leads to saturation feeding with less actual food present (lower values of I), a useful result. And in a natural setting, the effects of patchiness are even more important, for food is naturally patchy! As Ivlev says, the effects of patchiness "... are of very essential significance for natural conditions, in which it is never possible to observe an absolutely even distribution of food material for any animals whatsoever" [Ivlev 1961, 24].

The problem of patchiness in the food supply has been studied by other researchers. See [Pielou 1969] for an example.

References

- Gause, G. F. [1934] 1954. *The Struggle for Existence*, Reprint. New York: Dover Press. A readable and fundamental book in mathematical modeling in biology. Most of it is accessible to undergraduate mathematics students. The name is pronounced "Gow-suh."
- Ivlev, Victor Sergeivich. 1961. *Experimental Ecology of the Feeding of Fishes*. Translated from Russian by D. Scott, New Haven: Yale University Press.
- Lotka, A. J. 1925. *Elements of Physical Biology*. Baltimore. A ground-breaking book in mathematical modeling.
- Newling, Bruce. 1969. The spatial variation of urban population densities. *Geographical Review* 59: 242-252.
- Pielou, E. C. 1969. *An Introduction to Mathematical Ecology*. New York: Wiley.
- Roberts, Fred S. 1983. Efficiency of energy use on obtaining food, I: Humans. Chap. 10 in *Life Science Models*, edited by Helen Marcus-Roberts and Maynard Thompson, 250-285. Modules in Applied Mathematics, vol. 4. New York: Springer-Verlag.
- Roberts, Fred S., and Helen Marcus-Roberts. 1983. Efficiency of energy use in obtaining food, II: Animals. Chap. 11 in *Life Science Modules*, edited by Helen Marcus-Roberts and Maynard Thompson, 286-349. Modules in Applied Mathematics, vol. 4. New York: Springer-Verlag.
- Schoener, T. W. 1969. Models of optimal size for solitary predators. *American Naturalist* 103: 273-313.
- Smith, David A. et. al. 1985. Three observations on a theme. *Mathematics Magazine* 58 (3): 146-150.
- Smith, John Maynard. 1974. *Models in Ecology*. Cambridge: Cambridge University Press.
- Volterra, V. 1933. *Leçons sur la théorie mathématique de lutte pour la vie*. Paris.
- Volterra, V., and U. d'Ancona. 1935. *Les associations biologiques au point de la vue mathématique*. Paris.
- Watt, Kenneth E. F. 1968. *Ecology and Resource Management*. New York: McGraw-Hill.
- Weatherley, A. H. 1972. *Growth and Ecology of Fish Populations*. New York: Academic Press.

Journal articles in biology are easily located by use of the journals *Biological Abstracts* and *BioResearch Index*.

Exercises

1. Sometimes we have experimental data and two or more potential models. The task is to decide which model the data “belong to.” There are advanced methods for doing this, but in this exercise we’ll point out elementary graphical methods. (The more sophisticated methods give more certain decisions, however.)

- a. The solution of $dr/dI = a(R - r)$, $r(0) = 0$ was given in Section 3 as $r = R(1 - \exp(-aI))$. Algebraically equivalent

$$\ln\left(1 - \frac{r}{R}\right) = -aI. \quad (\text{E1})$$

Let’s examine a graphical method of deciding whether data fit this model.

First, pick specific values of a and R and obtain a table of half a dozen (I, r) data pairs from $r = R(1 - \exp(-aI))$. That’s “data set 1.”

From data set 1, form data sets 2 and 3: set 2 has data pairs $(I, \ln(1 - r/R))$ and set 3 has pairs $(I, 1 - r/R)$. Graph set 2 on ordinary (Cartesian) graph paper; graph set 3 on semilogarithmic graph paper.

In each case, the data should fall on a straight line through the origin. This is true for set 2 because you have graphed (x, y) pairs that obey $y = -ax$, as (E1) really says. As for set 3, the spacing of the lines on semilog graph paper is such that, although you have labeled the axes with (x, y) values, you are really graphing $(x, \log y)$ pairs. Thus, set 3 falls on the straight lines as claimed. *The point:* If experimental (I, r) data pairs lead to $(I, 1 - r/R)$ data pairs that very nearly fall on a straight line through the origin on semi-logarithmic graph paper, the fit to $r = R(1 - \exp(-aI))$ is good.

- b. The so-called logistic model will be our alternative to the model of Exercise 1a above: it is

$$\frac{dr}{dI} = kr(R - r)$$

and has many applications. Use integration by partial fractions to show that the solution is

$$r(I) = \frac{RB e^{kRI}}{1 + B e^{kRI}} = \frac{RB}{e^{-kRI} + B}, \quad (\text{E2})$$

where B is the constant of integration. Show that $r(I) \rightarrow R$ as $I \rightarrow \infty$: the logistic model has the same asymptotic behavior as Ivlev’s model in Exercise 1a.

- c. Algebraically rearrange (E2) into

$$\ln\left(\frac{R}{r} - 1\right) = -\ln B - kRI.$$

If we draw (I, r) data pairs from (E2) and form $(I, \ln(R/r - 1))$ data pairs from them, how will the latter graph on Cartesian graph paper? How can we use semilog paper to decide whether experimental (I, r) data pairs obey the logistic model?

2. a. Show that $dr/dI = a(R - r)$ and $d(R - r)/dI = -a(R - r)$ are equivalent equations.
- b. Let $Z = R - r$. Show that the solution of $dZ/dI = -aZ$ is $Z = Be^{-aI}$, i.e., classical exponential decay (B is the constant of integration). Deduce the solution formula of $dr/dI = a(R - r)$, $r(0) = 0$ from the formula for Z .
- c. If you are familiar with the notion of “half-life” for the equation $dy/dt = -ky$, whose solution is $y = Be^{-kt}$, interpret “half-life” for the function $r(I) = R(1 - e^{-aI})$.
3. Deduce eq. (6) as the solution of eq. (5).
4. A method of combining the models of r depending separately on I and p into one model without use of calculus: Recall that r , in terms of the patchiness alone is, from Section 4,

$$r(p) = (R - \rho)(1 - e^{-bp}) + \rho;$$

and that ρ , being the value of r when $p = 0$, is really $\rho = R(1 - e^{-aI})$ by Section 3. Substitute this p into $r(p)$ to get $r(I, p) = R(1 - e^{-aI - bp})$, the same model we got from the multivariable chain rule. [The very special conditions here let us avoid the chain rule, but it is still the fundamental method needed to combine one-dimensional results and is usually the only available tool.]

5. The *level curves* of the surface $z = f(x, y)$ are the curves obtained by intersecting planes parallel to the xy plane ($z = \text{const}$) with the surface. We can calculate them by putting $z = k$, a constant, and solving for y in terms of x to get the points that are graphed at height k . For example, putting $z = k$ in the equation of a sphere centered at $(0, 0, 0)$,

$$x^2 + y^2 + z^2 = 16,$$

yields

$$x^2 + y^2 = 16 - k^2 = \text{constant},$$

which is a circle (the level curve) when $|k| \leq 4$, but is the null set for $|k| > 4$. These level curves are the parallels of latitude if the z -axis passes through the north pole and the equator is in the xy plane.

Your problem: show that the level curves of our solution surface

$$r = r(I, p) = R(1 - e^{-aI - bp})$$

are straight lines.

The differential equation $dy/dt = k(K - y)$, which we have used in Sections 3 and 4, has many other applications. We will mention two:

6. If t represents time, y is the temperature of an object in degrees, and K is the temperature of the surrounding room (considered to be constant), then *Newton's*

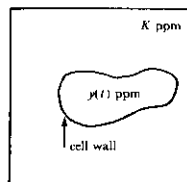
Law of Cooling is

$$\frac{dy}{dt} = k(K - y),$$

where k is a constant of proportionality that depends on the object. Your problem:

A cake is taken from a 350°F oven at 2 P.M. and is allowed to cool in a 70° room. At 2:30 P.M., its temperature is 150°F. It can be sampled when its temperature reaches 95°. When do we eat?

7. During osmosis, nutrients dissolved in the surrounding fluid bath move through the wall of a cell. If the surrounding fluid contains a concentration of K parts per million nutrients, considered to be constant, and $y(t)$ is the concentration inside the cell as a function of time, then *Fick's Law of Osmosis* is $dy/dt = k(K - y)$, where k is again a constant of proportionality. Problem: Suppose $y = K/2$ when $t = 0$ and use $k = 1$. At what times will $y(t) = \frac{3}{4}K$? $\frac{7}{8}K$? $\frac{15}{16}K$? Extrapolate based on the pattern that has appeared.



Exercises A1 through A8 require an understanding of the least-squares material in Appendix A.

- A1. The least squares fit of $B = mP$ in Example 1 of Appendix A involves $S(m) = \sum(mP_i - B_i)^2$.

- Show that $S(m) = m^2 \sum P_i^2 - 2m \sum P_i B_i + \sum B_i^2$.
- What sort of function is $S: m \rightarrow S(m)$? Notice that $\sum P_i^2$, $\sum P_i B_i$, and $\sum B_i^2$ are simply constants.
- Calculate dS/dm from the form of $S(m)$ in (a), and the point m_0 where $dS/dm = 0$. (For this simple case, this is an easier way to get m_0 .)
- Establish carefully that the local minimum $(m_0, S(m_0))$ is an absolute minimum for all real m .

- A2. In Example 2 of Appendix A, Section 2, we showed that

$$S(a, b) = \sum (a + bx_i - y_i)^2$$

has only one critical point among all (a, b) in the plane. Show that this point must be a local minimum, for which the criterion is

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0 \quad \text{and} \quad \frac{\partial^2 S}{\partial a^2} \cdot \frac{\partial^2 S}{\partial b^2} - \left(\frac{\partial^2 S}{\partial a \partial b} \right)^2 > 0.$$

- A3. a.** Algebraically show that $y = ae^{bx} \Leftrightarrow \ln y = \ln a + bx$. This says that variables x and y will obey the exponential model if and only if x and $\ln y$ have a straight line relationship.
- b.** You are given (x_i, y_i) data, $i = 1, 2, \dots, N$ that fit $y = ae^{bx}$ roughly and you want a, b . If we convert the (x_i, y_i) data into $(x_i, z_i) = (x_i, \ln y_i)$ how can we proceed to select a and b ? You want to exploit the straight-line relationship of x and z by use of Example 2, Appendix A, Section 2.
- A4. a.** Show that $y = ax^b \Leftrightarrow \ln y = \ln a + b \ln x$. This says that x and y are power-function related if and only if $\ln x$ and $\ln y$ are straight-line-related.
- b.** Given data (x_i, y_i) that follow $y = ax^b$, how can we exploit (a) to select the “best” values of a, b ?
- A5. a.** We want to select a, b , and c by the least squares method to fit a quadratic (parabola) $y = a + bx + cx^2$ to given data (x_i, y_i) . What function $S(a, b, c)$ should we seek to minimize? What three linear equations does this lead us to simultaneously solve?
- b.** Your answer to (a) should be a natural extension of the two linear equations we finally reached in Example 2, Appendix A, Section 2. What do you suppose the four equations will be that fit $y = a + bx + cx^2 + dx^3$ to (x_i, y_i) data? Calculate with $S(a, b, c, d)$ to confirm your conjecture.
- A6. a.** How can we select a and b to fit $y = a \sin x + b \cos x$ to data (x_i, y_i) ?
- b.** How will the results of (a) allow you to fit $y = a \sin(x + b)$ to data (x_i, y_i) ? [We can similarly fit $y = a \cos(x + b)$.]
- A7.** We have data triples (x, y, z) and wish to choose parameters, a, b, c by the method of least squares to fit the model

$$z = ae^{bx}y^c$$

to the data. Set up $S(a, b, c)$ as the appropriate sum of squares that will lead to linear equations for a, b, c . Hint: Take logs on both sides of the formula for z . This model has applications. For example, Schoener [1969] uses the formula to model $z =$ “time used by a predator in eating its prey” in terms of $x =$ “body size of prey” and $y =$ “body size of predator.”

- A8.** We have data pairs (x, y) and wish to choose parameters a, b, c by the method of least squares to fit the model

$$y = e^{a+bx-cx^2}$$

to the data. Set up $S(a, b, c)$ so that linear equations for a, b, c result. (Hint: Compare problems A5 and A7.) An application: Newling [1969] uses this formula to successfully model $y =$ “residential population density (in people per unit area)” in terms of $x =$ “distance outward from the center of the city.”

Solutions to the Exercises

- 1c.** A straight line of slope $-kR$ and intercept $-\ln(B)$ should appear on Cartesian paper. Data pairs $(I, R/r - 1)$ yield the same straight line on semilog paper.

2b. From $Z = B \exp(-aI)$ and $Z = R - r$ it follows that $r = R - B \exp(-aI)$. Use of the initial value completes the problem.

c. $r - r = R \exp(-aI)$ has the half-life phenomenon of being cut in half each time I grows by $\ln(2)/a$. Thus, the gap between r and its asymptote decays with the half-life pattern.

3. The solution of $dr/dp = b(R - r)$ is $r = R - C \exp(-bp)$, as done in Section 3. The initial value $r(0) = \rho$ leads to $C = R - \rho$, thus

$$\begin{aligned} r &= R - (R - \rho) \exp(-bp) \\ &= \rho + (R - \rho) - (R - \rho) \exp(-bp) \\ &= \rho + (R - \rho)(1 - \exp(-bp)). \end{aligned}$$

4. The algebra is straightforward.

5. From $k = R(1 - \exp(-aI - bp))$ we get $-aI - bp = \ln(1 - k/R) = \text{constant}$, which is a family of straight lines in the (I, p) plane.

6. Given: (1) $dy/dt = k(K - y)$, $K = 70$;

and (2) $y = 350$ at $t = 2$ and $y = 150$ at $t = 2.5$;

Wanted: t such that $y = 95$. From (1) deduce $y = 70 - C \exp(-kt)$ for some constants C, k . From (2) next get $k = 2.5$ and $C = -42000$ (not many significant digits should be kept). Now from $y = 70 + 42000 \exp(-2.5t)$ and $y = 95$ get $t = 2.97$; in hours, minutes, and seconds, this is 2:58:12 P.M.

7. The reader should discover the half-life pattern of this differential equation by exploration here. The conditions $dy/dt = k(K - y)$, $k = 1$, and $y(0) = K/2$ lead to $y = K - \frac{1}{2}K \exp(-t)$. Then $y(t) = 3K/4, 7K/8$ and $15K/16$ occur at $t = \ln(2), 2 \ln(2)$, and $3 \ln(2)$ respectively. Extrapolation leads to the half-life pattern for $K - y(t)$, which may also be derived analytically.

A1b. A parabola, opening upward.

d. If $m = m_1$ were another candidate for the minimum, then we would have some closed interval containing both m_0 and m_1 in its interior. A standard calculus theorem (stating that the absolute min must occur at a horizontal tangent, endpoint, or point where the derivative fails to exist, when the domain is a closed interval and the function is continuous) may then be applied.

A2. The criterion for a local min requires that we show

$$(\sum x)^2 < N \sum (x^2).$$

(We abbreviate a typical x_i as x here.) Both sums run $i = 1, 2, \dots, N$. From $0 \leq (x_i - x_j)^2$ [and we get the $<$ case if $x_i \neq x_j$] we can deduce that

$$2x_i x_j \leq x_i^2 + x_j^2, \quad \text{with } < \text{ if } x_i \neq x_j.$$

Summing, we get the desired result with \langle as long as any two of the x_i are different.

A3b. From data pairs (x, y) , form a new data set $(x, z) = (x, \ln(y))$. Fit the (x, z) data to a straight line, getting A and B such that $\ln(y) = A + Bx$ is the model. Equivalently, $y = \exp(A)\exp(Bx) = a \exp(bx)$ is the exponential model for the original (x, y) data, where $a = \exp(A)$ and $B = b$ should be used.

A4b. From (x, y) data pairs form $(\ln(x), \ln(y))$ pairs. Do a straight-line fit on the latter set of data, getting A and B such that $\ln(y) = A + B \ln(x)$. Then the model for the original (x, y) data is $y = \exp(A)\exp(B \ln(x)) = ax$ where $a = \exp(A)$ and $b = B$ are the correct parameters.

A5a. $S(a, b, c) = \sum(a + bx_i + cx_i^2 - y_i)^2$ leads to these linear equations:

$$\begin{pmatrix} N & \Sigma x & \Sigma(x^2) \\ \Sigma x & \Sigma(x^2) & \Sigma(x^3) \\ \Sigma(x^2) & \Sigma(x^3) & \Sigma(x^4) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \end{pmatrix}$$

Here Σx means Σx_i , $\Sigma x^2 y$ means $\Sigma x_i^2 y_i$, etc., with all sums $i = 1, 2, \dots, N$.

b. $S(a, b, c, d) = \sum(a + bx_i + cx_i^2 + dx_i^3 - y_i)^2$ leads to these equations, given in matrix form:

$$\begin{pmatrix} N & \Sigma x & \Sigma(x^2) & \Sigma(x^3) \\ \Sigma x & \Sigma(x^2) & \Sigma(x^3) & \Sigma(x^4) \\ \Sigma(x^2) & \Sigma(x^3) & \Sigma(x^4) & \Sigma(x^5) \\ \Sigma(x^3) & \Sigma(x^4) & \Sigma(x^5) & \Sigma(x^6) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \\ \Sigma x^3 y \end{pmatrix}$$

A6a. Use $S(a, b) = \sum(a \sin(x_i) + b \cos(x_i) - y_i)^2$; these linear equations result:

$$\begin{pmatrix} \Sigma \sin^2(x_i) & \Sigma \sin(x_i)\cos(x_i) \\ \Sigma \sin(x_i)\cos(x_i) & \Sigma \cos^2(x_i) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Sigma y_i \sin(x_i) \\ \Sigma y_i \cos(x_i) \end{pmatrix}$$

b. $y = a \sin(x + b) = a(\sin(x)\cos(b) + \cos(x)\sin(b))$
 $= [a \cos(b)]\sin(x) + [a \sin(b)]\cos(x)$

Perform the least squares method of Exercise A6a on the (x, y) data, getting $A = a \cos(b)$ and $B = a \sin(b)$. The needed a, b are then

$$a = \sqrt{A^2 + B^2} \quad \text{and} \quad b = \arctan(B/A).$$

A7. $z = ae^{bx}y^c \Leftrightarrow \ln(z) = \ln(a) + bx + c \ln(y)$.

From (x, y, z) data triples, form new data triples $(X, Y, Z) = (x, \ln(y), \ln(z))$ and fit the model $Z = A + BX + CY$ to that new data. This may be done via

$S(A, B, C) = \sum (A + BX_i + CY_i - Z_i)^2$, which leads us to get A, B, C from:

$$\begin{pmatrix} N & \sum X_i & \sum Y_i \\ \sum X_i & \sum (X_i)^2 & \sum X_i Y_i \\ \sum Y_i & \sum X_i Y_i & \sum (Y_i)^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \sum Z_i \\ \sum X_i Z_i \\ \sum Y_i Z_i \end{pmatrix}$$

The parameters for Schoener's model are then $a = \exp(A)$, $b = B$, and $c = C$.

A8. $y = \exp(a + bx - cx^2) \Leftrightarrow \ln(y) = a + bx - cx^2$.

From the (x, y) data pairs, form $(x, \ln(y))$ pairs and do the quadratic fit of Exercise A5 to the new data. The parameters that emerge are the needed values of $a, b, -c$ for Newling's model.

