ESTIMATION OF THE RELATIVE FISHING POWER OF INDIVIDUAL SHIPS

D. S. Robson*

BU-133-M

May, 1961

Introduction

The use of commercial catch statistics in constructing indices of fish abundance such as catch per unit of effort requires that the unit of effort be well defined and constant through time. Since the total commercial catch for a year is made up of the individual annual catches of a number of ships of varying types and sizes working several different kinds of gear this requirement for a standard unit of effort raises difficulties. Beverton and Holt (1) consider this problem in connection with the analysis of catch statistics from the plaice fishery of the North Sea, and point up the need for a statistically efficient method of estimating the relative fishing power of each vessel so that the actual effort of the vessel can be transformed into standard units on a scale comparable to that of all other vessels.

The problem is neatly illustrated by Beverton and Holt with an example in which the total catch and hours of effort is known for 6 fishing trips involving 3 different vessels and 3 different locations in time and space (Table 1). They suggest that one of the trips, say ship A at location 1, be

Table 1. The catch rates for 6 different fishing trips

	Location		
Ship	I	II	III
A	A_1	A_2	--
B	B_1	B_2	B_3
C	--	--	C_3

arbitrarily selected as a standard for comparison and the fishing power of the other two vessels then be expressed relative to the fishing power of vessel A. Thus, four possible estimates of the fishing power of vessel C are listed as

* Biometrics Unit, Plant Breeding Department, Cornell University

(i) $\frac{B_1}{A_1} \times \frac{C_3}{B_3}$ (ii) $\frac{B_2}{A_2} \times \frac{C_3}{B_3}$ (iii) $\frac{1}{2}(\frac{B_1}{A_1} + \frac{B_2}{A_2}) \times \frac{C_3}{B_3}$ (iv) $\frac{B_1+B_2}{A_1+A_2} \times \frac{C_3}{B_3}$

with the suggestion that (iv) might be "as good as any". We shall show that for this problem the estimator

$$(v) \quad \frac{C_3}{B_3}\sqrt{\frac{B_1B_2}{A_1A_2}}$$

is likely to be more efficient than any of the others, and shall indicate the general method by which such efficient estimators are constructed.

A Statistical Model for Catch Rates

The model upon which the 4 estimators of Beverton and Holt are based is the multiplicative model for a two-way classification without interaction. Thus, the model asserts that one location is expected to yield a catch rate which is a fixed percentage higher than another location for every vessel and, equivalently, that one vessel is expected to achieve a catch rate which is a fixed percentage higher than that of another vessel at every location. So, denoting P_i as the power factor of the i'th vessel and Q_j as the power factor of the j'th location we obtain the model shown in Table 2.

Table 2. Mulitplicative model for the catch rates of 3 vessels at 3 locations

	Location		
Ship	1	2	3
1	$CP_1Q_1E_{11}$	$CP_1Q_2E_{12}$	$CP_1Q_3E_{13}$
2	$CP_2Q_1E_{21}$	$CP_2Q_2E_{22}$	$CP_2Q_3E_{23}$
3	$CP_3Q_1E_{31}$	$CP_3Q_2E_{32}$	$CP_3Q_3E_{33}$

where C is a constant and E is a random variable having an expected value of unity. If the trip of ship number 1 (ship A) to location number 1 is to be taken as the standard for comparison then all other power factors P and Q are to be expressed as a fraction of P_1 and Q_1; hence, in this instance we would set $P_1=Q_1=1$ and obtain for this example the model shown in Table 3. Putting the errors E all equal to 1 then reveals the intuitive basis of the estimators (i)-(v)

Table 3. Multiplicative model for the 6 trips with ship A and location I taken as standard

	Location		
Ship	I	II	III
A	$A_1=CE_{11}$	$A_2=CQ_2E_{12}$	--
B	$B_1=CP_2E_{21}$	$B_2=CP_2Q_2E_{22}$	$B_3=CP_2Q_3E_{23}$
C	--	--	$C_3=CP_3Q_3E_{33}$

given earlier.

Empirical evidence in support of this multiplicative model for catch statistics of plaice was presented by Beverton and Holt; they found that when fishing power statistics were classified according to tonnage and method of propulsion of the vessels the distribution of errors within these classes was log normal -- that is, on the logarithmic scale the within-class distribution of power factors was normal with constant variance. Their subsequent analyses were therefore performed on the log scale, and on the basis of this finding it is now apparent that a more efficient method of estimating power factors would have been to transform to the logarithmic scale at the beginning and compute the least squares or maximum likelihood estimators of fishing power. On this scale the multiplicative model of Table 2 becomes additive as shown in Table 4, where lower case letters are used to denote

Table 4. Additive model for catch rates on the logarithmic scale

	Location		
Ship	1	2	3
1	$c+p_1+q_1+e_{11}$	$c+p_1+q_2+e_{12}$	$c+p_1+q_3+e_{13}$
2	$c+p_2+q_1+e_{21}$	$c+p_2+q_2+e_{22}$	$c+p_2+q_3+e_{23}$
3	$c+p_3+q_1+e_{31}$	$c+p_3+q_2+e_{32}$	$c+p_3+q_3+e_{33}$

logarithms ($x=\log X$). According to the evidence presented by Beverton and Holt, the errors e are normally distributed with zero mean and constant variance.

Relative fishing power may be defined for this model in almost any way that is convenient, the only formal requirement being that the definition impose linear restraints on the p's and q's. The most conventional linear restrictions are

$$p_1+p_2+p_3=0 , \qquad q_1+q_2+q_3=0$$

which amount to using the row and column means as standares for comparison; under these restrictions the constant c in the model then represents the average log catch rate for the set of 3 ships and 3 locations. The procedure suggested earlier of choosing one ship and one location as the standards for comparison is perhaps equally convenient, though seemingly more arbitrary. Such a choice as ship number 1 and location number 1 is then equivalent to imposing the linear restrictions

$$p_1=\log P_1=\log 1=0, \qquad q_1=\log Q_1=\log 1=0$$

The choice of linear restraints is irrelevant except from the standpoint of computational convenience, for using the statistically most efficient method of estimation will produce the same <u>relative</u> fishing powers regardless of the choice; that is, the difference between two estimates $\hat{p}_1-\hat{p}_2$ will be independent of the choice of linear restrictions. This point will be illustrated with the example involving six fishing trips.

<u>Least Squares of Maximum Likelihood Estimation of Fishing Power</u>

The method of estimating the parameters of an additive tow-factor model with missing cells appears in standard textbooks on statistical methodology (see, for example, Steel and Torrie (1960) p.289) usually under the name of "the method of fitting constants." This procedure is an application of the least squares method of multiple regression which, under the normality assumptions mentioned earlier, also yields maximum likelihood estimates. The method is algebraically simple, but is computationally fairly tedious, involving matrix inversion, though with electronic computers this, too, becomes a relatively simple operation.

In the textbook treatment of this topic the emphasis is ordinarily placed upon hypothesis testing rather than point estimation, and is presented under the general heading of "analysis of variance." For estimation purposes it is perhaps more convenient to regard this problem as a special case of the general multiple regression problem expressed by the model

$$Y_i=\beta_0X_{0i}+\beta_1X_{1i}+\cdots+\beta_kX_{ki}+\epsilon_i$$

where the expected value of an observation Y depends linearly upon the levels of k+1 factors, and each observed Y_i is obtained at a different set of levels $X_{0i},X_{1i},\cdots,X_{ki}$ of these factors. The X's are known constants, the β's are unknown and to be estimated, and the ϵ's are independent identically distributed errors. If n observations $Y_1,\cdots,Y_n$ are available then the best unbiased estimator of $\beta'=(\beta_0,\beta_1,\cdots\beta_k)$ is, in matrix notation, $\hat{\beta}=(X'X)^{-1}X'Y$; if the ϵ's are normally distributed then this is also the maximum likelihood estimator and is statistically efficient. We illustrate this estimation procedure first with the data of Table 3 transformed to logarithms as shown in Table 5.

Table 5. Additive model for the log catch rates of 6 fishing trips with ship A and location 1 taken as standard

Ship	Location 1	Location 2	Location 3
A	$a_1=c+e_{11}$	$a_2=c+q_2+e_{12}$	--
B	$b_1=c+p_2+e_{21}$	$b_2=c+p_2+q_2+e_{22}$	$b_3=c+p_2+q_3+e_{23}$
C	--	--	$c_3=c+p_3+q_3+e_{33}$

The role of the β parameters is now taken by $\beta'=(c,p_2,p_3,q_2,q_3)$ and the X matrix of coefficients of these parameters is then given by Table 6. Taking the sums of squares and crossproducts of the columns of this matrix, we then obtain the matrix X'X shown in Table 7; and inverting this, an operation which would require only a second or two on a high speed computer, we obtain the matrix $(X'X)^{-1}$ shown in Table 7. Finally, the product of $(X'X)^{-1}$ with the

Table 6. Matrix X of coefficients of the parameters in a linear additive model

Observation	Parameter c	p_2	p_3	q_2	q_3
a_1	1	0	0	0	0
a_2	1	0	0	1	0
b_1	1	1	0	0	0
b_2	1	1	0	1	0
b_3	1	1	0	0	1
c_3	1	0	1	0	1

Table 7. Matrix X'X of sums of squares and crossproducts of the columns of X and the inverse $(X'X)^{-1}$

Column	Column c	p_2	p_3	q_2	q_3
c	6	3	1	2	2
p_2	3	3	0	1	1
p_3	1	0	1	0	1
q_2	2	1	0	2	0
q_3	2	1	1	0	2

The inverse of X'X

	c	p_2	p_3	q_2	q_3
c	3/4	-1/2	-1/2	-1/2	-1/4
p_2	-1/2	1	1	0	-1/2
p_3	-1/2	1	3	0	-3/2
q_2	-1/2	0	0	1	1/2
q_3	-1/4	-1/2	-3/2	1/2	7/4

crossproducts

$$X'Y = \begin{bmatrix} a_1+a_2+b_1+b_2+b_3+c_3 \\ b_1+b_2+b_3 \\ c_3 \\ a_2+b_2 \\ b_3+c_3 \end{bmatrix}$$

gives the estimates

log scale	original scale
$\hat{c} = \frac{1}{4}(3a_1+a_2+b_1-b_2)$	
$\hat{p}_2 = \frac{1}{2}(b_1+b_2-a_1-a_2)$	$\hat{P}_2 = \sqrt{B_1B_2/A_1A_2}$
$\hat{p}_3 = \hat{p}_2-b_3+c_3$	$\hat{P}_3 = \hat{P}_2C_3/B_3$
$\hat{q}_2 = \frac{1}{2}(b_1-b_2-a_1+a_2)$	
$\hat{q}_3 = b_3-\frac{1}{4}(a_1-a_2+3b_1+b_2)$	

The variance of these estimates on the logarithmic scale is σ_ϵ^2 times the diagonal elements of $(X'X)^{-1}$; thus, $\sigma^2_{\hat{c}}=3\sigma^2_\epsilon/4$, $\sigma^2_{\hat{p}_2}=\sigma^2_\epsilon$, $\sigma^2_{\hat{p}_3}=3\sigma^2_\epsilon$, etc. Covariances between estimates are likewise computed as σ_ϵ^2 times the corresponding element of $(X'X)^{-1}$; thus, $\sigma_{\hat{p}_2,\hat{p}_3}=\sigma^2_\epsilon$, $\sigma_{\hat{p}_2,\hat{q}_2}=0$, $\sigma_{\hat{p}_3,\hat{q}_3}=-3\sigma^2_\epsilon/2$, etc., and so $\sigma^2_{\hat{p}_2-\hat{p}_3}=\sigma^2_{\hat{p}_2}+\sigma^2_{\hat{p}_3}-2\sigma_{\hat{p}_2,\hat{p}_3}=2\sigma^2_\epsilon$.

If instead of the restriction $p_1=q_1=0$ we impose the restriction $p_1+p_2+p_3=q_1+q_2+q_3=0$ then essentially the same results will be obtained. The X matrix for this case is shown in Table 8, along with the inverse of X'X, from which we obtain the estimates in Table 8.

Table 8. The matrices X and $(X'X)^{-1}$ for the case $p_1+p_2+p_3=q_1+q_2+q_3=0$

X

	Parameter				
Observation	c	p_1	p_2	q_1	q_2
a_1	1	1	0	1	0
a_2	1	1	0	0	1
b_1	1	0	1	1	0
b_2	1	0	1	0	1
b_3	1	0	1	-1	-1
c_3	1	-1	-1	-1	-1

$(X'X)^{-1}$

	Parameter				
Parameter	c	p_1	p_2	q_1	q_2
c	2/9	-1/9	-1/9	1/18	1/18
p_1	-1/9	2/3	0	-2/9	-2/9
p_2	-1/9	0	1/3	-1/18	-1/18
q_1	1/18	-2/9	-1/18	5/12	-1/12
q_2	1/18	-2/9	-1/18	-1/12	5/12

Estimates

$\hat{c} = \frac{1}{6}[a_1+a_2+b_1+b_2+2c_3]$ $\qquad \sigma^2_{\hat{c}} = \frac{1}{6}\sigma^2_\epsilon$

$\hat{p}_1 = \frac{1}{3}[a_1+a_2-b_1-b_2+b_3-c_3]$ $\qquad \sigma^2_{\hat{p}_1} = \frac{2}{3}\sigma^2_\epsilon$

$\hat{p}_2 = \frac{1}{3}[\frac{1}{2}(b_1+b_2-a_1-a_2)+b_3-c_3]$ $\qquad \sigma^2_{\hat{p}_2} = \frac{1}{3}\sigma^2_\epsilon$

$\hat{q}_1 = \frac{1}{12}[3a_1-3a_2+5b_1-b_2-4b_3]$ $\qquad \sigma^2_{\hat{q}_1} = \frac{5}{12}\sigma^2_\epsilon$

$\hat{q}_2 = \frac{1}{12}[3a_2-3a_1-b_1+5b_2-4b_3]$ $\qquad \sigma^2_{\hat{q}_2} = \frac{5}{12}\sigma^2_\epsilon$

$\hat{p}_3 = -\hat{p}_1-\hat{p}_2$ $\qquad \sigma^2_{\hat{p}_3} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} + 2\sigma_{\hat{p}_1\hat{p}_2} = \sigma^2_\epsilon$

$\hat{q}_3 = -\hat{q}_2-\hat{q}_3$ $\qquad \sigma^2_{\hat{q}_3} = \sigma^2_{\hat{q}_1} + \sigma^2_{\hat{q}_2} + 2\sigma_{\hat{q}_1,\hat{q}_2} = \frac{2}{3}\sigma^2_\epsilon$

We note that the fishing power of ship B relative to ship A is again estimated by

$$\hat{p}_2-\hat{p}_1 = \frac{1}{2}(b_1+b_2-a_1-a_2) \qquad \sigma^2_{\hat{p}_2-\hat{p}_1}=\sigma^2_{\hat{p}_1}+\sigma^2_{\hat{p}_2}-2\sigma_{\hat{p}_1,\hat{p}_2}=\sigma^2_\epsilon$$

so

$$\frac{\hat{P}_2}{\hat{P}_1} = \sqrt{\frac{B_1B_2}{A_1A_2}}$$

as before, and

$$\hat{p}_3-\hat{p}_1 = -2\hat{p}_1-\hat{p}_2 = \hat{p}_2-\hat{p}_1-b_3+c_3 \ , \quad \sigma^2_{\hat{p}_3-\hat{p}_1}=4\sigma^2_{\hat{p}_1}+\sigma^2_{\hat{p}_2}+4\sigma_{\hat{p}_1,\hat{p}_2}=3\sigma^2_\epsilon$$

so, as before,

$$\frac{\hat{P}_3}{\hat{P}_1} = \frac{C_3}{B_3}\sqrt{\frac{B_1B_2}{A_1A_2}}$$

The residual variance σ^2_ϵ is estimated in the general regression problem by $(Y'Y-\hat{\beta}X'Y)/(n-k-1)$. In the present example n=6, and k+1=5 parameters are being estimated so the residual sum of squares

$$(a_1^2+a_2^2+b_1^2+b_2^2+b_3^2+c_3^2)-\hat{c}(a_1+a_2+b_1+b_2+b_3+c_3)\ -\hat{p}_1(a_1+a_2-c_3)\ -\hat{p}_2(b_1+b_2+b_3-c_3)$$

$$-\hat{q}_1(a_1+b_1-b_3-c_3)\ -\hat{q}_2(a_2+b_2-b_3-c_3)$$

has only 1 degree of freedom. An analagous form for the residual applies to the earlier analysis with $p_1=q_1=0$ and gives the same value for $\hat{\sigma}^2_\epsilon$. For purposes of hypothesis testing it is pertinent that $\hat{\sigma}^2_\epsilon$ is statistically independent of the other estimates. Thus, for example, under the hypothesis that $p_1=p_3$ the ratio $(\hat{p}_1-\hat{p}_3)/\sqrt{3\hat{\sigma}^2_\epsilon}$ is distributed as Student's t with 1 degree of freedom.

Modifications of the Model

Beverton and Holt list a number of factors such as tonnage, class, design, age, and skipper which might explain the variation in fishing power of the different vessels, suggesting that a classification of vessels into various groups should be introduced into the model to determine the effect of such factors.

The collection of fishing locations in time and space might also be further classified, at least according to time and geographic location, to determine the effects of these factors on catch rates; since the same principles apply to the subclassification of ships and locations, however, it will suffice to consider only the former.

In the North Sea plaice fishery a major factor affecting catch rate was the means of propulsion of the vessel; Beverton and Holt distinguished two classes for this factor, the steam trawler and the diesel powered motor trawler. To illustrate the modification of the model required to incorporate this factor we enlarge our earlier example to include 5 ships, say 3 steam trawlers and 2 motor trawlers as shown in Table 9. The earlier model of Tables 2 and 4 would still suffice for the 3 steam trawlers and an entirely similar model should hold

Table 9. The catch rates of two types of vessels at three locations

Ship	Type	Location I	II	III
A	Steam	A_1	A_2	--
B	Steam	B_1	B_2	B_3
C	Steam	--	--	C_3
F	Motor	F_1	--	F_3
G	Motor	--	G_2	--

for the 2 motor trawlers; that is, the additive model of Table 4 should now be extended to include

Ship	Type	Location I	II	III
F	Motor	$c+p_4+q_1+e_{41}$	$c+p_4+q_2+e_{42}$	$c+p_4+q_3+e_{43}$
G	Motor	$c+p_5+q_1+e_{51}$	$c+p_5+q_2+e_{52}$	$c+p_5+q_3+e_{53}$

where, under the conventional linear restrictions,

$$p_4+p_5=0 , \qquad q_4+q_5=0 .$$

The constant c, representing the average log catch rate, should however be different for steam and motor trawlers if method of propulaion actually does affect catch rate. We therefore identify these two constants as c_1 and c_2, respectively, and taking $c=(c_1+c_2)/2$ as our standard for comparison we let

$$c_1=c+(c_1-c)=c+d_1 \ , \quad c_2=c+(c_2-c)=c+d_2 \ , \quad d_1+d_2=0$$

The additive model for the log catch rates of Table 9 which incorporates this d-effect due to method of propulstion is shown in Table 10; with the 5 linear restrictions which have been imposed there are only 7 independent parameters in this table, and their estimation proceeds as before by the methods of multiple regression. A test of the significance of the effect of method of propulsion would then be obtained from the regression analysis as a t- or F-test of the hypothesis that the "regression coefficient" d_1 is equal to zero.

Table 10. Additive model for log catch rate of two types of ships

Ship	Type	Location I	Location II	Location III	Log Tonnage
A	Steam	$c+d_1+p_1+q_1+e_{11}$	$c+d_1+p_1+q_2+e_{12}$	--	t_1
B	Steam	$c+d_1+p_2+q_1+e_{21}$	$c+d_1+p_2+q_2+e_{22}$	$c+d_1+p_2+q_3+e_{23}$	t_2
C	Steam	--	--	$c+d_1+p_3+q_3+e_{33}$	t_3
F	Motor	$c+d_2+p_4+q_1+e_{41}$	--	$c+d_2+p_4+q_3+e_{43}$	t_4
G	Motor	--	$c+d_2+p_5+q_2+e_{52}$	--	t_5

Another factor which proved to have a significant effect upon catch rate was the size of the ship as measured by its tonnage; in fact, it was found that the power factor P was directly proportional to tonnage, on the average, with different proportionality factors for steam and motor trawlers. The constants d_1 and d_2 of the above model would represent these two proportionality factors on the log scale, and incorporation into the model of the assumption that the power factor P_i is proportional to the tonnage T_i would then consist of replacing the p-parameters by the log tonnage deviates,

$$p_1=(t_1-\bar{t}_S) \qquad p_4=(t_4-\bar{t}_M)$$
$$p_2=(t_2-\bar{t}_S) \qquad p_5=(t_5-\bar{t}_M)$$
$$p_3=(t_3-\bar{t}_S)$$

where $t_i = \log T_i$ and

$$\bar{t}_S=(t_1+t_2+t_3)/3 \qquad \bar{t}_M=(t_4+t_5)/2$$

The procedure for fitting this modified model would be to deduct the log tonnage deviate $t-\bar{t}$ from each of the observed log catch rates of that ship, eliminate all p's from the model and proceed with a multiple regression analysis. The difference between the residual sum of squares from this analysis and the residual sum of squares from the analysis of Table 10 represents the reduction in residual sum of squares which is attained due to fitting the p's instead of simply assuming that the p's are equal to the corresponding known constants $t-\bar{t}$. The degrees of freedom in this difference of residuals is equal to the number of ships minus the number of types, or 5-2=3 in this case, and the mean square obtained by dividing by degrees of freedom may then be tested for significance against the residual mean square from the analysis of Table 10. The entire procedure is illustrated by a numerical example in the next section.

Numerical Illustration

The preceding analysis is illustrated here with an artificial set of data for Table 10; these data, shown in Table 11, were generated by assigning arbitrary values to the c, d, q and t parameters of Table 10, and taking the p parameters approximately equal to the log tonnage deviates $t-\bar{t}$ within each type of vessel. Observed log catch rates (Y) were then constructed by combining the parameters in the manner indicated in Table 10 and adding to each a random normal deviate e.

The first phase of the analysis consists of setting out the coefficients of the unknown parameters in matrix form (Table 12) and computing crossproducts among these coefficients and between the coefficients and the observations. This matrix of crossproducts is then inverted by standard methods and multiplied by the vector of crossproducts between coefficients and observations to obtain

Table 11. Log catch rates of five ships at three locations

Ship	Type	Y=log catch rate Location I	II	III	log Tonnage	Y*=log catch rate[1] adjusted Location I	II	III
A	Steam	.020	.120	--	1.954	.207	.307	--
B	Steam	.503	.463	.238	2.322	.322	.282	.057
C	Steam	--	--	.251	2.146	--	--	.246
		Mean log tonnage			2.141			
F	Motor	.188	--	.142	1.778	.356	--	.310
G	Motor	--	.544	--	2.114	--	.376	--
		Mean log tonnage			1.946			

[1]Adjusted log catch rate = log catch rate - log tonnage + mean log tonnage

Table 12. Coefficients of the unknown parameters (Table 10) and their cross-products with log catch rates

Catch rate Y	Y*	Parameter c	d_1	p_1	p_2	q_1	q_2	p_4
.020	.207	1	1	1	0	1	0	0
.120	.307	1	1	1	0	0	1	0
.503	.322	1	1	0	1	1	0	0
.463	.282	1	1	0	1	0	1	0
.238	.057	1	1	0	1	-1	-1	0
.251	.246	1	1	-1	-1	-1	-1	0
.188	.356	1	-1	0	0	1	0	1
.142	.310	1	-1	0	0	-1	-1	1
.544	.376	1	-1	0	0	0	1	-1
ΣXY		2.469	.721	-.111	.953	.080	.496	-.214
	ΣXY*	2.463	.379			.272	.352	

Crossproducts of coefficients of the parameters

X'X

Parameter	c	d_1	p_1	p_2	q_1	q_2	p_4
c	9	3	1	2	0	0	1
d_1	3	9	1	2	0	0	-1
p_1	1	1	3	1	2	2	0
p_2	2	2	1	4	1	1	0
q_1	0	0	2	1	6	3	0
q_2	0	0	2	1	3	6	-2
p_4	1	-1	0	0	0	-2	3

the parameter estimates. Thus, in Table 13, the estimate $\hat{p}_1=-.111$ is obtained as the sum of crossproducts of the inverse elements in the p_1 column times the elements of the ΣXY column,

$$\hat{p}_1 = [(-10)(2.469)+(-34)(.721)+\cdots+(-72)(-.213)]/540=-.111$$

Goodness of fit of the model may be measured by the ratio

$$R^2 = \frac{\text{Residual S.S. after fitting (c)} - \text{Residual S.S. after fitting } (c,d_1,p_1,p_2,q_1,q_2,p_4)}{\text{Residual S.S. after fitting c}}$$

The residual sum of squares (S.S.) after fitting the constant c (the mean) is simply the corrected sum of squares of the 9 observations,

$$\text{Res. S.S. after fitting (c)} = \sum_1^9 Y_i^2 - \frac{1}{9}(\Sigma Y)^2$$
$$= .953267 - \frac{1}{9}(2.469)^2 = .275938 \qquad (\text{d.f.}=8)$$

and the residual sum of squares after fitting all seven parameters is, from Table 13,

$$.953267-[(.297)(2.469)+(-.037)(.721)+\cdots+(-.126)(-.214)]$$
$$=.953267-.939339=.013928 \qquad (\text{d.f.}=2)$$

giving

$$R^2 = \frac{.275938-.013928}{.275938} = \frac{.262010}{.275938} = .95$$

Thus, 95 percent of the variance among the 9 observations is accounted for by the six parameters d_1, p_1, p_2, q_1, q_2 and p_4; this, however, is not statistically significant when tested by the F-test

$$F = \frac{\text{Mean Square due to fitting } d_1,p_1,p_2,q_1,q_2,p_4}{\text{Residual Mean Square after fitting } d_1,p_1,p_2,q_1,q_2,p_4}$$

$$= \frac{.262010/6}{.013928/2} = \frac{.043668}{.006964} = 6.2$$

because of the small number of degrees of freedom in the residual.

Table 13. Solution to the multiple regression problem of Table 12

Inverse matrix of crossproducts of coefficients (x540)

$(X'X)^{-1}$

Parameter	c	d_1	p_1	p_2	q_1	q_2	p_4	ΣXY	Estimate
c	80	-25	-10	-25	15	-15	-45	2.469	.297
d_1	-25	86	-34	-31	- 3	39	63	.721	-.037
p_1	-10	-34	296	-16	-48	-96	-72	-.111	-.247
p_2	-25	-31	-16	176	-12	-24	-18	.953	.142
q_1	15	- 3	-48	-12	144	-72	-54	.080	.030
q_2	-15	39	-96	-24	-72	216	162	.496	.084
p_4	-45	63	-72	-18	-54	162	324	-.214	-.126

Inverse of deleted matrix of crossproducts
of coefficients (x72)

Parameter	c	d_1	q_1	q_2	ΣXY	Estimate
c	9	-3	0	0	2.469	.279
d_1	-3	9	0	0	.721	-.013
q_1	0	0	16	-8	.080	-.037
q_2	0	0	-8	16	.496	.101
ΣXY*	2.463	.379	.272	.352		
Estimate	.292	-.055	.021	.048		

Analogous methods may be used to test the significance of more specific features of the model; for example, to determine the fraction of the total variance which is due specifically to variation of fishing power among ships of like method of propulsion we compute

$$\frac{\text{Res.S.S. after fitting } (c,d_1,q_1,q_2) - \text{Res.S.S. after fitting } (c,d_1,p_1,p_2,q_1,q_2,p_4)}{\text{Res.S.S. after fitting } c}$$

This computation involves another matrix inversion; we simply drop the p's from the model and so reduce the coefficient matrix by deleting the blocks indicated in Table 12. Estimates of c, d_1, q_1 and q_2 obtained from the inverse of the deleted matrix are given in Table 13; thus

$$\hat{c} = [9(2.469)-3(.721)+0(.080)+0(.496)]/72 = .279$$

and the residual S.S. after fitting only c, d_1, q_1 and q_2 is then

$$.953267-[(.279)(2.469)+(-.013)(.721)+(-.037)(.080)+(.101)(.496)] = .227363$$

with 9-4=5 degrees of freedom. The fraction of the total variance which is due specifically to the p-parameters is therefore

$$\frac{.227363-.013928}{.275938} = \frac{.213435}{.275938} = .77$$

and the significance of this may be tested by

$$F = \frac{\text{M.S. due to } (p_1,p_2,p_4)}{\text{Res.M.S.}} = \frac{.213435/3}{.013928/2} = 10.2$$

Even with 2 degrees of freedom in the residual, the variation in the p's is detected at the 10 percent significance level.

The difference between the average log catch rate of steam and motor trawlers is measured by the parameter d_1, which may be tested for significance in the above manner. For a single parameter, however, the test procedure illustrated above simplifies to an F-test of the form

$$F = \frac{(\hat{d}_1)^2}{\frac{86}{540}\text{ Res. M.S.}} = \frac{(-.037)^2}{\frac{86}{540}(.006964)} = 1.23$$

which is non-significant. The fraction 86/540 is the d_1 diagonal element of the inverse matrix of Table 13.

Finally, we illustrate the procedure for testing the hypothesis that within types of vessels, the fishing power of a ship is proportional to its tonnage. On the log scale, this is equivalent to the hypothesis that within types of vessels, $p_i=t_i-\bar{t}$. This hypothesis is tested by replacing p_i by $t_i-\bar{t}$ in the model and comparing the resulting residual M.S. with the residual M.S. of .006964 obtained with no restrictions on p_1, p_2 and p_4. Since the deviates $t_i-\bar{t}$ are known constants then the replacement of p_i by $t_i-\bar{t}$ in the model is equivalent to subtracting $t_i-\bar{t}$ from the observed log catch rates of ship number i; the resulting adjusted catch rates are shown in Table 11 and denoted by Y^*. With the p-parameters thus deducted from the model, the matrix of parameter coefficients becomes the deleted matrix considered previously, and the computations follow the same pattern but with Y replaced by Y^*. The parameter estimates so obtained (Table 13) then give

$$\text{Res.S.S. after fitting } (c^*, d_1^*, p_1=-.187,\ p_2=.181,\ q_1^*, q_2^*, p_4=-.168)$$

$$=\sum_1^9 Y^{*2}-c^*(2.463)-d_1^*(.379)-q_1^*(.272)-q_2^*(.352)$$

$$=.748283-(.292)(2.463)-(-.055)(.379)-(.021)(.272)-(.048)(.352)$$

$$=.027123$$

Permitting p_1, p_2 and p_4 to vary arbitrarily in the model thus reduces the residual sum of squares by only

$$.027123-.013928=.013195$$

and this sum of squares with 3 degrees of freedom is not significant when tested against the residual for the unrestricted model,

$$F = \frac{.013195/3}{.013928/2} = .63 \quad .$$

Another way of expressing this result is that while the best fitting p-parameters accounted for 77 percent of the variation in log catch rate, fixing the p-parameters by making them equal to the log tonnage deviates reduced this percentage only to .72 percent,

$$\frac{.213435-.013195}{.375938} = .72$$

and the reduction was not significant.

References

1) Beverton, R. J. H. and S. J. Holt. On the Dynamics of Eiploited Fish Populations. Fisheries Investigations, Series II, Volume XIX, 1957.

2) Steel, R. G. D. and J. H. Torrie. Principles and Procedures of Statistics. McGraw-Hill Book Company, Inc., 1960.