Malacological Review, 1971, 4: 17-20

THE STEPWISE MULTIPLE REGRESSION METHOD FOR SELECTION OF VARIABLES FOR PREDICTING THE SHELL WEIGHT OF FRESHWATER MUSSELS

Billy G. Isom

ABSTRACT

The stepwise multiple regression method was used for selection of variables for predicting the shell weight of the freshwater mussel *Pleurobema cordatum* (Rafinesque, 1820). Analyses showed that regression of weight on width of the bivalves gave a correlation coefficient (R^2) of 0.9240. The regression of weight on length R^2 of 0.9591 was 0.0351 more than width, but the easier measurement of width indicates it would be the best single measure for predicting shell weight of *P. cordatum*.

The objective of this study was to determine which of several independent variables were related to weight of mussels and to rate these variables in order of their importance for prediction of shell weight. Selection of a single variable of the set that could be measured in the field and yield a high degree of confidence was the primary goal of the experiment. Valve thickness constitutes the most desirable morphometric quality in selection of shells for commercial use, while hardness is the most desirable physical quality.

Data for 4 variables were available for analyses. These variables were weight (wt), height (h), width (w), and length (l) measurements on the freshwater mussel *Pleurobema cordatum* (Rafinesque, 1820). Weight was designated as the dependent variable in these analyses.

The data analyzed were collected by the author in 1963-1964 below Guntersville Dam, Tennessee River (Isom, 1969). The 32 sets of data were taken in 468 one-third-square-yard bottom samples. These samples were taken randomly from one-tenth-square-mile plots. *Pleurobema cordatum* is a species that has been and continues to be of great value to the freshwater bivalve industry in the Tennessee Valley. Animals are taken with commercial gear (brails) and sold to buyers who in turn sell the shells to Japanese representatives of the cultured pearl industry. The shells are cut and turned into round beads that are used as nuclei for cultured pearls.

A stepwise multiple regression program was used to find and evaluate the variables with the best multiple correlation (R^2) for prediction of the Y variable (wt) of *Pleurobema cordatum*. A computer program was used to compute all cross-products of the independent variables (w, h, and l) as shown in Table 1. Computations were made on the IBM 360-67 computer at the University of Michigan, using a canned program, BMDO2R, stepwise regression. Table 2 is a list of the values for the dependent variable (Y), the independent variables (X_i) and the cross-products of the independent variables. A computer program was also used to punch data cards containing the information noted in Table 2.

The multiple linear regression model is $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \; . \; . \; . \; . \; + \beta_k X_k + \varepsilon$ where α and β_i are constants and where the errors ε are distributed independently with zero mean and variance σ^2 (Snedecor & Cochran, 1967).

R^2 is the fraction of the sum of squares of deviations of Y from its mean that is attributable to the regression while $(1-R^2)$ is the fraction not associated with

TABLE 1. Variable name and number

Weight (gm) - wt	1
Width (mm) - w	2
Height (mm) - h	3
Length (mm) - l	4
Width x height	5
Width x length	6
Height x length	7
Width x height x length	8

the regression. One hundred times R^2 equals the percent sum of squares of deviations of Y from its mean that is attributable to the regression. If R^2 is a small fraction of unity, variation in Y is unexplained due to random variation or independent variables not considered in the regression.

The computer analysis starts with the regression of Y on the X_i variable shown to give the greatest reduction in the sum of squares of Y. The variable that made the next greatest contribution was then selected and the process continued until an additional contribution b_i^2/c_{ii} was too small to satisfy rules for inclusion (F-level 0.01 included, F-level 0.005 deleted) with a tolerance level of 0.001. The regression of weight on each single variable, on each pair of variables, and on the single triplet variable was calculated. The program also gave means and standard deviations.

Analyses indicate that weight of the shells of *Pleurobema cordatum* are highly correlated with all of the independent variables. The computer analysis established that variable 8 (Table 2) and weight R^2 was 0.9920. This is an unusually high correlation coefficient for biological material as compared with many of those noted in Snedecor & Cochran (1967). Weight on variable 6 R^2 of 0.9883 was the 2nd greatest contribution to reduction in the residual sum of squares of Y. R^2 for length and weight was 0.9591. Further analysis of the data involved only the regression of weight on width which revealed an R^2 of 0.9240.

Knowledge of the ecology of the species and of its commercial uses indicates that thickness of the individual valves of *Pleurobema cordatum* is the most desirable quality; the thicker the valves the larger and better are the nuclei that can be made from the shells. Therefore, thickness of the individual valves would be a desirable measurement, but such a measurement can be made only after the animal is removed. An alternative method would be to simply measure the greatest width of the intact bivalves, if high correlation existed between width and weight. As already noted, weight was more highly correlated with length than width; however, measurement on length in bivalves entails some ambiguity since the anterio-posterior plane is not often easily recognized, whereas width is easily measured. Since the difference in R^2 between length and weight and width and weight is only 0.0351, width would be the more desirable measure to make in the "real world." This would result in only a minor error in prediction of weight, but at less cost.

sum of squares of
ression. If R^2 is a
ndom variation or

on the X_i variable
of Y. The variable
ed and the process
all to satisfy rules
) with a tolerance
iable, on each pair
The program also

rdatum are highly
iter analysis esta-
his is an unusually
ared with many of
ible 6 R^2 of 0.9883
sum of squares of
of the data involved
0.9240.

cial uses indicates
datum is the most
tter are the nuclei
e individual valves
t can be made only
to simply measure
on existed between
ly correlated with
alves entails some
easily recognized,
R^2 between length
the more desirable
only a minor error

TABLE 2. Original input data and data calculated by computer program

Case No.	Wt (1)	W (2)	H (3)	L (4)	W. H. (5)	W. L. (6)	H. L. (7)	W. H. L. (8)
1	9.700	18.6	27.0	28.4	502.20	528.20	766.80	14262.50
2	9.400	18.6	25.4	28.5	472.40	530.10	723.90	13464.50
3	15.800	22.2	31.3	32.9	694.90	730.40	1029.80	22860.90
4	12.300	19.9	29.3	33.0	583.10	656.70	966.90	19241.30
5	16.200	22.2	30.3	34.0	672.70	754.80	1030.20	22870.40
6	13.900	20.5	31.6	34.6	647.80	709.30	1093.40	22413.90
7	17.600	23.0	33.9	36.1	779.70	830.30	1223.80	28147.10
8	21.600	24.6	35.1	36.3	863.50	893.00	1274.10	31343.60
9	19.600	24.2	32.4	36.9	784.10	893.00	1195.60	28932.50
10	20.800	23.2	34.4	36.9	798.10	856.10	1269.40	29449.10
11	18.300	23.6	33.6	37.6	793.00	887.40	1263.40	29815.30
12	26.300	24.1	39.1	42.8	942.30	1031.50	1673.50	40330.80
13	33.100	27.4	42.0	46.0	1150.80	1260.40	1932.00	52936.80
14	37.100	27.8	45.6	48.9	1267.70	1359.40	2229.80	61989.50
15	45.900	29.3	49.4	50.5	1447.40	1479.60	2494.70	73094.60
16	43.400	32.9	43.5	51.5	1431.10	1694.30	2240.20	73704.20
17	49.100	32.2	45.5	52.9	1465.10	1703.40	2406.90	77503.80
18	52.200	32.5	49.2	56.8	1599.00	1846.00	2794.60	90823.10
19	62.300	35.1	52.4	57.2	1839.20	2007.70	2997.30	105204.40
20	66.800	35.5	53.4	58.8	1895.70	2087.40	3139.90	111467.10
21	79.700	37.7	53.4	61.7	2013.20	2326.10	3294.80	124213.10
22	83.800	40.7	60.8	62.5	2474.60	2543.70	3800.00	154659.90
23	86.400	38.7	63.0	65.0	2438.10	2515.50	4095.00	158476.40
24	109.100	45.0	60.9	67.8	2740.50	3051.00	4129.00	185805.80
25	120.700	43.1	65.5	69.7	2823.00	3004.10	4565.30	196766.50
26	96.500	38.7	61.8	70.0	2391.70	2709.00	4326.00	167416.10
27	111.200	41.2	64.0	72.9	2636.80	3003.50	4665.60	192222.70
28	117.200	40.0	62.3	74.4	2492.00	2976.00	4635.10	185404.70
29	140.500	46.1	66.3	74.6	3056.40	3439.10	4946.00	228009.50
30	135.600	44.9	68.9	77.7	3093.60	3488.70	5353.50	240373.40
31	155.000	44.0	76.0	87.4	3344.00	3845.60	6642.40	292265.60
32	151.600	44.4	68.8	88.0	3054.70	3907.20	6054.40	268815.20

CONCLUSIONS

The stepwise multiple regression technique showed that volume (w × h × l) is the best measure for prediction of weight (Y) of *Pleurobema cordatum*.

The best single measure for prediction of weight was length.

Further analysis and knowledge of the ecology and qualitatively desirable characteristics of shells used in the pearl culture industry indicate width of the intact bivalve is a good predictor of weight. Width denotes a good correlation with weight, with an R^2 value of 0.9240, and would be an easy and cheap measure for the prediction of weight of a population of *Pleurobema cordatum*.

The stepwise multiple regression method of analysis would seem adaptable to study clinal characteristics often noted in *Pleurobema cordatum* and other bivalves as well as gastropods. However, correlation analysis (Cvancara, 1963), partial correlation analysis, or the discriminant function (Snedecor & Cochran, 1967) may be better tools for clinal studies.

Knowledge of the different morphometries of various clines and their relation to weight would be a valuable tool for determining the value of mussel shells from a given locality for use in the culture of pearls or for classification studies.

Other variables, such as water quality, could be introduced. Calcium, hardness and current velocity, for example, could be correlated with morphometric data. Addition of such variables introduces the possibility of added correlation among X_i values which makes determination of the most important variable more difficult. Since, for example, calcium constitutes better than 99% of bivalve shells, calcium content of the water should be correlated with all X_i variables (shell morphometry).

Variable 8 (Table 2) and weight correlation coefficient (R^2) of 0.992 indicates 99.2% of the variation in Y can be attributed to its linear regression on variable 8. Variable 6 (Table 2) and weight correlation coefficient of 0.9883 indicates 98.8%, variable 4 (Table 2) and weight, 95.9%; and variable 2 and weight, 92.4% of the variation of Y that can be attributed to its linear regression on these X variables.

LITERATURE CITED

CVANCARA, A. M. 1963. Clines in three species of *Lampsilis* (Pelecypoda: Unionidae). *Malacologia*, 1(2): 215-225.

ISOM, B. G. 1969. The mussel resource of the Tennessee River. *Malacologia*, 7(2-3): 397-425.

SNEDECOR, G. W. & COCHRAN, W. G. 1967. *Statistical Methods*. Iowa State Univ. Press, 593 p.

B. G. ISOM

Environmental Biology Branch
Division of Environmental Research and Development
Tennessee Valley Authority
Muscle Shoals, Alabama 35660
U. S. A.

MALACOLOGICAL REVIEW, P.O.Box 801, Whitmore Lake, Michigan 48189, U.S.A.