68. Ecological Determinants of Population*

M. A. Baumhoff

The topic of the present paper is a general method for describing and analyzing population determinants. By the word determinant I mean a factor which is definable in some terms other than population, and the variation of which tends to be associated with a variation in population. If the nature of the association between the determinant and the population can be explicitly defined we may be able to express the association as a mathematical function. In so doing we acquire the precision and elegance of the mathematical art for purposes of prediction, description and understanding.

Suppose we begin with a single determinant and examine the nature of the relationship we wish to establish. Let the determinant be fishing-miles and the populations be those of certain California Athabascan groups (see Table 1).

Table 1

Group	Population	Fishing-miles
Wailaki	1,656	23
Pitch Wailaki	1,104	15
Mattole	1,200	38.5
Lolangkok Sinkyone	2,076	63
Hupa	1,475	39
Whilkut	2,588	70

The California Athabascans are a fishing people so we may expect the quantity of fish resources to affect in some way their population, although we would be surprised if there were a one-to-one correspondence. Fishing-miles, as the term is used here, refers to statute miles along salmon streams. There is evidence of a strong correlation between this statistic and total fish resources; Rostlund, in his paper on North American fishery (1952, p. 17), says that the aboriginal fishermen of

* Read at the annual meeting of the Kroeber Anthropological Society, May 18, 1957.

California in no case approached over-fishing. If this is so there must have been ample fish left uncaught even in the smallest salmon streams so that a mile of one stream was about equivalent to a mile of another.

The population figures given in Table 1 were derived from ethnographic data--village counts, house counts, and the like. These population estimates are about three times larger than the comparable estimates given by Kroeber in his Handbook of California Indians (1925, p. 883) and may therefore be subject to some question. I am prepared to defend the estimates on the basis of evidence available to me which Kroeber did not have in 1925. In any case the population estimates are used here merely for purposes of illustration and should at least be adequate for that.

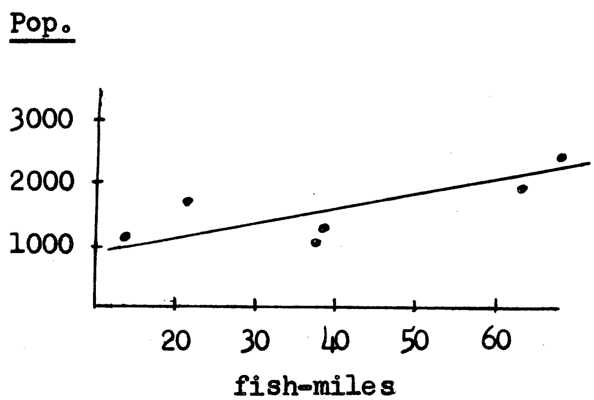

Figure 1

Population--fish-miles

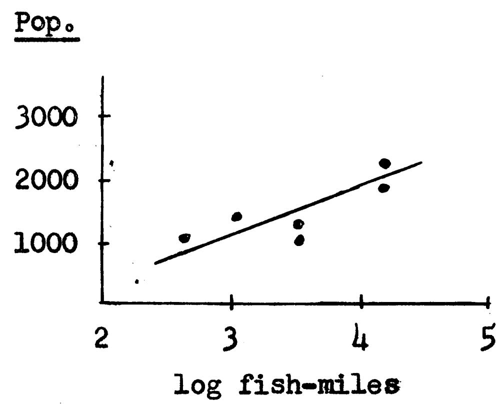

Figure 2

Population--log fish-miles

If we plot population against fish-miles we obtain the scatter diagram shown in Figure 1. A glance at the diagram shows that increases in population tend to be associated with increases in fish-miles but the association is not in any sense precise--there is no obvious mathematical function whose curve would go through the points on the diagram. Suppose we try to transform our data so as to make the relationships more obvious. One transformation often used in a situation of this kind is the logarithmic function. Figure 2 shows population plotted against the log of fish-miles. This is perhaps slightly better than before although the points still show considerable scatter about the straight line fitted to them. There may be some other transformation which would meet the situation better but I have not been able to find it. The whole subject of transformations of this kind is in need of an exhaustive analysis both as to which transformation is appropriate and as to the meaning of the transformation once it has been accepted.

If we conclude that the logarithmic transformation is adequate, our model is shown in equation (1).

$$\text{population} = m + b(\log \text{fish-miles}) \quad (1)$$

where m and b are constants to be determined from the data. In the present instance the equation is,

$$\text{population} = -1 + .75(\log \text{fish-miles}) \quad (2)$$

determined by the method of least squares from the data in Table 1. This is the equation of the line shown in Figure 2. The equation accounts for the general trend of the data but does not account for the scattering of points on either side of the line. In order to deal with this we add another term, obtaining equation (3),

$$\text{population} = -1 + .75(\log \text{fish-miles}) + E \quad (3)$$

where E is the so-called error term. E is not constant but fluctuates randomly in some sense. It is common in cases of this kind to assume that E has a normal probability distribution with an average value of zero and some constant but unknown standard deviation. That is to say E has the ordinary bell-shaped probability curve centered at zero. If this assumption is correct we are in a position to predict, within certain specified limits, the population of a group merely from the fishing-miles available to it. Or at least we can so predict if we feel the situation is similar to the one which produced the equation. For instance the Kato (another California Athabascan group) had 29 fishing miles available to them so their predicted population is 1523 ± 267 with 80% confidence. That is to say under the foregoing assumptions 80% of all groups with 29 fishing miles available would have populations somewhere between 1256 and 1790.

Now let us examine the assumptions once more. We have assumed that population is a linear function of log fishing-miles but with an added factor which fluctuates randomly. A further assumption was that the error term E was normally distributed with average value zero and with some constant but unknown standard deviation or variance. Why do we assume that E is normal, that it has the familar bell-shaped curve? To answer this we must enquire into the sources of error. Fundamentally the error stems from two factors:

(a) Measurement. The population figures given in Table 1 are subject to considerable error due to our inability to take a direct census. In almost all cases there is at least some difficulty in this respect. Also the fishing-miles cannot be measured with absolute accuracy.

(b) Multiplicity of determinants. It is clear that there are some things, and probably many, other than fishing-miles affecting the size of populations, and our predictions are less accurate to the degree that these things are not taken into account.

Since measurement errors and multiplicity of determinants are the factors underlying the observed variation it seems safe to conclude that the error term is normally distributed, for these are precisely the kinds of factors which ordinarily produce a normally distributed variable.

The other assumptions made about the error term were that its average is zero and its standard deviation is constant. It is reasonable to assume that the average or mean value is zero since the line was drawn in such a way that the points are scattered the same distance on either side of it. That the standard deviation is constant is not so clear but ordinarily the failure of this assumption will not radically alter the predictions if they are about the same order of magnitude as the data on which they are based.

If the analysis is thus far correct then the next problem is to reduce the standard deviation or variance so that the prediction will gain in precision, in other words, that the confidence intervals will be shorter. We would like to find a curve or line to which the points would cling more intimately. One approach would be to try for increased accuracy in measurement, but this is usually not feasible. As far as one can tell the measurements are made as accurately as possible already and the residual error is due to faulty data which cannot be rectified. However this factor should obviously be kept in mind as new information is gathered.

The other source of variation is multiplicity of determinants, a matter which can be more adequately dealt with. Suppose we have reason to believe that the magnitude of the area occupied by a group has a considerable effect on population size and we wish to take this factor into account. If we call (F) the effect due to fish resources and (A) the effect due to area then our model will look like equation (4)

$$\text{population} = m + (F) + (A) + E \qquad (4)$$

where m is the general mean and E is the usual error term. If this were the complete model equation there would be no difficulty in adding any number of additional terms in the same manner but already there are further problems. The way equation (4) is set up an increase in the area-effect produces the same increase in population regardless of the magnitude of the fishing resources. But it is a clear possibility that a fishing people whose population pressure is maximum with respect to fish resources might not respond at all to an increase in territory. In other words the effect of (A) might depend on the level of (F) and the effect of (F) might depend on the level of (A). To take care of this possibility we add another factor to the model and obtain equation (5)

$$\text{population} = m + (F) + (A) + (FA) + E \qquad (5)$$

The (FA) factor is called the interaction term and complicates matters exceedingly. If we wish to consider a third factor, say (D) for deer resources, we obtain the still more complicated model shown in equation (6)

$$\text{population} = m + (F) + (A) + (D) + (FA) + (FD) + (AD) + (FAD) + E \qquad (6)$$

The total number of possible terms in the model equation increases exponentially with the number of factors considered. We might hope that some of the interaction terms would be negligible and could therefore be ignored in the model equation. Things would be greatly simplified if this were true but obviously this is a matter to be tested rather than assumed.

The model equations can assume a variety of forms, depending on the situation, each form having its own difficulties. Now suppose that we have completed enough investigation to conclude that equation (7) is a satisfacmodel for some group.

$$\text{population} = m + a(F) + b(A) + c(FA) + E \qquad (7)$$

The constants a, b, and c indicate the relative importance of each determinant. As noted previously the model enables us to predict populations where they are unknown and if this is true it should also be possible to predict past populations, at least to a limited extent. In order to predict the population of an ancient group we would have to know the values of F* and A, that is fishing-miles and area, and also to have some estimate of the constants m, a, b, and c.

Obtaining the pertinent information about fishing-miles and area is essentially a problem in the investigation of prehistoric ecology. This is a problem which is not easily solved but there has recently been good progress along these lines, notably by certain of the European archaeologists (see Clark, 1952) and by Heizer and Cook at the University of California (see Bibliography). In order to estimate the constants m, a, b, and c in equation (7) it is necessary to have some idea of the structure of the economy. These constants are essentially weight factors and to know them would be equivalent to knowing the relative importance of each determinant in the economy. Archaeological data often yields this kind of information, at least in a general way. In this respect it would be quite useful to obtain the model equations for areas of known population and then compare the equations to the archaeological information from the same area. Again the work of Cook and Heizer, and of Cook and Treganza, on detailed midden analysis seems to be the most fruitful approach, at least for the local California situation.

None of the problems outlined here is unsolvable, I think, and their solution will put a rather neat tool in our hands. One fact stands out strongly in the present analysis and that is that we are forced to rely heavily on the deductive method. In the experimental disciplines it is feasible to make a number of sweeping generalizations rather casually,

---

* Note that "F" in equation (7) does not, as in equation (4), refer to <u>effect</u> due to fish resources.

confident that failure of the assumptions will be easily detected in the results of experimentation. With anthropology such is not the case; experiment is virtually non-existent and data available for testing by means of comparison are rather limited. For this reason it becomes all the more necessary for us to subject our assumptions to the most detailed scrutiny and to be extremely careful in our logic.

Bibliography

Clark, J. G. D.
1952 Prehistoric Europe: the Economic Basis. Methuen, London.

Cook, S. F.
1946 A Reconsideration of Shellmounds with Respect to Population and Nutrition. American Antiquity, Vol. 12, pp. 51-53.

1950 Physical Analysis as a Method of Investigating Prehistoric Habitation Sites. UCAS-R No. 7, pp. 2-5.

Cook, S. F. and R. F. Heizer
1951 The Physical Analysis of Nine Indian Mounds of the Lower Sacramento Valley. UC-PAAE Vol. 40, pp. 281-312.

Cook, S. F. and A. E. Treganza
1947 The Quantitative Investigation of Aboriginal Sites: Comparative Analysis of Two California Indian Mounds. American Antiquity, Vol. 13, pp. 135-141.

1950 The Quantitative Investigation of Indian Mounds. UC-PAAE Vol. 40, pp. 223-262.

Kroeber, A. L.
1925 Handbook of the Indians of California. Bur. of Amer. Ethnol. Bull. 78.

Rostlund, Erhard
1952 Freshwater Fish and Fishing in Native North America. Univ. of Calif. Publs. in Geography, Vol. 9.

Addenda

Following are bibliographic items referring to estimates of native population of California.

Baumhoff, M. A.
n.d. California Athabascan Groups. UC-AR Vol. 16, No. 5 (in press).

Additional information has made necessary a new estimate of population for these groups. My conclusions agree with Cook's (1956) in more than doubling Kroeber's estimate for the area.

Cook, S. F.
1943 The Conflict Between the California Indian and White Civilization: I. Univ. of Calif. Publs., Ibero-Americana, No. 21.

Cook re-evaluates Kroeber's (1925) data and concludes that the estimates should be raised by about 7 per cent.

---------- 1955 The Aboriginal Population of the San Joaquin Valley, California. UC-AR Vol. 16, No. 2.

------ 1956 The Aboriginal Population of the North Coast of California. UC-AR Vol. 16, No. 3.

------ 1957 The Aboriginal Population of Alameda and Contra Costa Counties, California. UC-AR Vol. 16, No. 4.

With the three papers above Cook has gone relatively far along into a complete re-estimate of California population. Cook has at hand much more ethnographic and especially historical data than were available to Kroeber. So far Cook's estimates seem to be about double those of Kroeber.

Cook, S. F. and A. E. Treganza
1950 Cited above.

Cook and Treganza deal here with the problem of deriving prehistoric population figures from archaeological data.

Heizer, R. F. and R. J. Squier
1953 Excavations at Site Nap-32 in July, 1951. In Archaeology of the Napa Region (R. F. Heizer, ed.). UC-AR Vol. 12, No. 3.

This paper includes an application of the methods of Cook and Treganza (1950).

Kroeber, A. L.
1925 Handbook of the Indians of California. Bur. of Amer. Ethnol., Bull. 78.

Kroeber here estimates aboriginal population at 133,000. His figures are based on detailed historical and ethnographical information for each tribe.

------ 1934 Native American Population. AA, Vol. 36, pp. 1-25.

This paper contains essentially the same information as was later published in Kroeber's Cultural and Natural Areas of Native North America.

------ 1939 Cultural and Natural Areas of Native North America. UC-PAAE Vol. 38.

In this paper Kroeber retains the population figures first published in his Handbook.

Merriam, C. Hart
1905 The Indian Population of California. AA Vol. 7, pp. 594-606.

Merriam estimates aboriginal population at 260,000. He bases his estimate on the number of baptisms in the coastal Mission strip and a simple extrapolation from this.

Powers, Stephen
1877 Tribes of California. U.S. Geographical and Geological Survey of the Rocky Mountain Region. Contributions to North American Ethnology, Vol. III.

Powers estimates total Indian population of California at 705,000 (p. 416). This figure is based on an estimate of natural resources together with a generalization from Yurok census data.