CONCENTRATION CURVES AND HAVE-STATISTICS FOR ECOLOGICAL ANALYSIS OF DIVERSITY:

PART I: DOMINANCE AND EVENNESS IN REPRODUCTIVE SUCCESS

Dianne G. Goodwin
James W. Vaupel

October 1985
WP-85-72

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

The present paper is an application of ideas derived from human populations and social statistics to animal ecology. Badly needed by both disciplines is a bridging of the gap that separates human from animal ecology, and the Goodwin-Vaupel research contributes strongly to this.

Both animal and human ecology have worked from averages in the past, and disregarded distributions. We talk of an average of 2 children surviving to maturity being needed for replacement in a low mortality society, and carry out calculations as though every woman must have two children. Such a condition is impossible, and the way in which actual populations depart from it is important. The fact that a substantial part of births are due to a relatively small fraction of women, and that everywhere many couples are sterile, considerably modifies some of the calculations made in demography.

But all this is provisional. The main significance of the present paper is in the linking together of two major disciplines.

Nathan Keyfitz
Leader
Population Program

PREFACE

Populations are heterogeneous. Much of demography and ecology (and of related fields in other disciplines) concerns different aspects of this heterogeneity. The two sexes are often distinguished, individuals are frequently classified by age, populations might be broken down by ethnic background (for humans) or species (in an ecological community); differences among individuals in reproductive success are of prime importance in evolutionary theory, as well as being of interest to demographers and others studying fertility. Numerous measures of variability, inequality, unevenness, diversity, and similarity have been developed to analyze heterogeneous populations and to make comparisons across populations or over time.

One of the focuses of the Population Program at IIASA in recent years has been the development and application of methods for the analysis of heterogeneous populations. This working paper is part of this tradition, although with a different emphasis than earlier (and ongoing) research on patterns of death and exit in heterogeneous populations. Here we focus on reproduction and seek to explain and demonstrate a method for capturing the pattern of heterogeneity in a population with regard to the concentration of reproduction among individuals. What proportion of the population has half the children? What proportion of the children does the most prolific quarter of the population have? We address questions like these by developing the notion of a concentration curve and some related summary statistics that we call have-statistics. Our examples are drawn from various animal species, but as noted below related analysis can shed light on human reproduction as well.

We see this paper as part of a larger set of papers dealing with several facets of "population concentration". Our hope is that the concept of population concentration will help population analysts in demography, ecology, and other disciplines study various aspects of population heterogeneity and its implications.

Five other papers on this topic are in various stages of completion. One concerns heterogeneity within populations of populations and focuses on the diversity and evenness of such communities with respect to their constituent populations of different species, ethnic groups, or whatever. Another paper reviews and evaluates a series of measures designed to summarize population concentration. The three other papers in progress all concern the concentration of reproduction in human populations. One of these papers focuses on changing patterns in the concentration of reproduction among U.S. females from 1917 to 1980. A second analyzes what might be called the age-concentration of reproduction: how many of the years of childbearing account for most of the children? Examples are drawn both from the United States and China and some implications for population stability are discussed. Finally, a paper that we are writing with Wolfgang Lutz compares the concentration of reproduction among women in over 40 countries and analyzes the relationship between the concentration of reproduction and such other factors as total fertility, education level, and rural vs. urban residence.

Our results so far indicate that reproduction is quite concentrated, for various animal species as well as for various human populations at different times and in different countries. Typically, something on the order of 10 to 25 percent of the females (or males) of one generation account for more than half of the offspring. In a cohort of red deer living on an island off the coast of Scotland, for instance, 23 percent of the females had half the offpsring surviving to one year of age. In the United States at 1980 reproduction rates, 25 percent of the women had half the children; at 1933 rates, 16 percent had half the children. Concentrations such as these have important evolutionary implications as well as some interesting policy implications.

ABSTRACT

Concentration curves and a set of summary statistics called have-statistics are useful in ecological analyses of dominance and evenness among individuals in reproductive success. This approach complements, but does not replace, approaches based on frequency distributions and standard summary statistics. Examples are drawn from studies of bullfrogs, red deer, elephant seals, sculpins, fruit flies, and rice weevils, as well as from some theoretical models similar to the Wright-Fisher model of evolutionary genetics.

KEYWORDS: diversity, dominance, evenness, inequality, concentration, reproductive success, Lorenz curves

ABOUT THE AUTHORS

Dianne G. Goodwin and James W. Vaupel are research scholars in the Population Program, led by Nathan Keyfitz, at the International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria. Goodwin formerly worked at the Institute of Animal Resource Ecology, University of British Columbia. Vaupel is also Professor of Public Affairs and Planning, University of Minnesota.

ACKNOWLEDGMENTS

We thank William Clark, Joel Cohen, Nathan Keyfitz, Warren Klein, Michael Stoto, Anatoli Yashin, and Peyton Young for their helpful comments and suggestions. In addition we are grateful to Eduard Löser and the library staff for their assistance, and Susan Stock and Linda Zwetkov for typing graphs and formatting the manuscript.

CONTENTS

In this paper, we illustrate the use of concentration curves and have-statistics as both conceptual and empirical tools for analyzing individual differences in reproductive success within a species, a topic of prime significance in the life sciences. We contrast concentration curves with frequency distributions, the usual method for presenting information about variation and diversity. Examples are drawn from studies of bullfrogs, red deer, elephant seals, sculpins, fruit flies, and rice weevils, as well as from some theoretical models related to the Wright-Fisher model of evolutionary genetics.

This paper is Part I of a trilogy of papers devoted to the use of concentration curves and have-statistics in ecological analyses. The second paper in the trilogy focuses on diversity among species; it also contains some discussion of applications to other ecological questions involving variation and inequality over space, time, and other dimensions. The third paper discusses various alternative summary measures of diversity and evenness and compares them with have-statistics and with the use of concentration curves. Our overall conclusion is that concentration curves and have-statistics are useful supplements to standard ecological methods for analyzing and summarizing diversity and evenness in reproductive success, species abundance, and other ecological topics.

Reproductive Success of Male Bullfrogs

Figure 1, which is adapted from Howard's (1983) important study of reproductive success in male bullfrogs, illustrates the use of a frequency distribution to depict diversity.[1] The figure shows the number of male bullfrogs at different levels of estimated seasonal reproductive success, defined as the number of hatchlings (hatched eggs) for each male in a population with mixed cohorts.

---

[1] All data used in the examples given in this paper are from the authors referenced. These authors rarely present data in raw form so it must be noted that while data were extracted as carefully as possible, errors are inevitable. The statistics should be viewed as illustrative rather than definitive and for purposes of substantive analysis original data sources should be used.

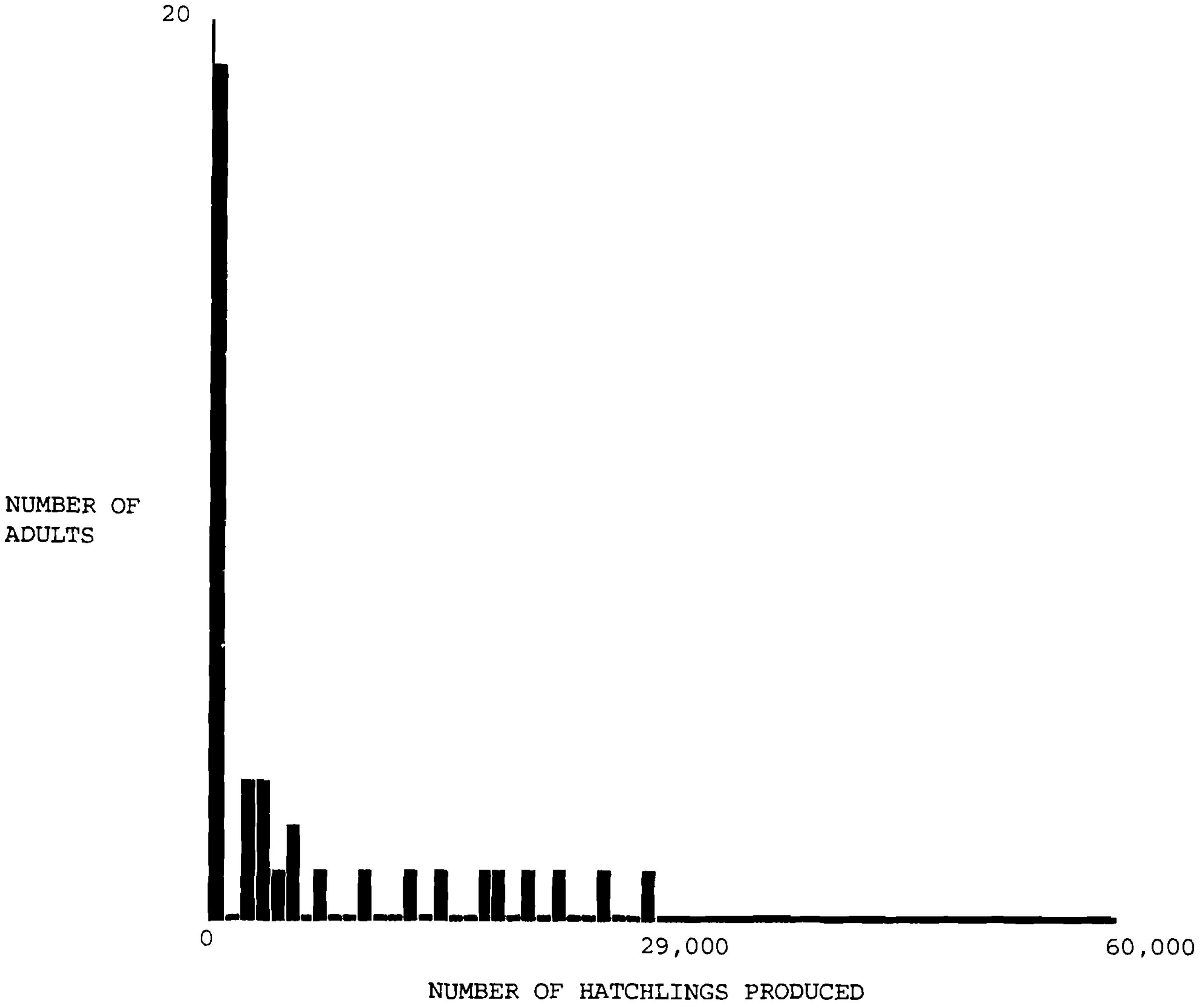


Figure 1: Estimated seasonal reproductive success in male bullfrogs-1976. (Redrawn from Howard, 1983)

Figure 2a, which we calculated based on the data in Figure 1, displays the seasonal concentration of reproduction among the adult male bullfrogs in this population. Both axes run from zero to one hundred percent: the horizontal axis gives the cumulative percentage of adult male bullfrogs, ranked by reproductive success, and the vertical axis gives the cumulative percentage of hatchlings. In Figure 2b the corresponding "have-statistics" are marked on the concentration curve:

- Half of the adult males had all the hatchlings and, correspondingly, half had no hatchlings;
- A quarter of the adults had 85 percent of the hatchlings;
- 11 percent of the adults (essentially 4 frogs out of the population of 38 adult frogs) had half the hatchlings;
- 5 percent of the adults (i.e., 2 frogs) had a quarter of the hatchlings.

The dotted diagonal line on the graph gives the concentration curve that would be observed if all the adult male frogs had exactly the same number of offspring: x percent of the frogs would have x percent of the hatchlings, for all values of x from zero to one hundred. Thus the area between the concentration curve and the 45 degree line provides a visual measure of the degree of concentration.

Table 1 indicates how Figures 2a and 2b were calculated from the data provided in Figure 1. Columns 1 and 2 in the table present the data from Figure 1 on number of adults and number of hatchlings. Note that the number of hatchlings are ranked from greatest number to least. The third column gives the percentage of adults, calculated by dividing the number in each category by the total number, and the fourth column gives the cumulative percentage. Similarly, the fifth and six columns give the percentage and the cumulative percentage of hatchlings. To calculate the concentration curves in Figures 2a and 2b, the numbers in the fourth column of the table were plotted against the numbers in the sixth column. The curve begins at the origin and ends at the point where 100 percent of the adults have 100 percent of the offspring.

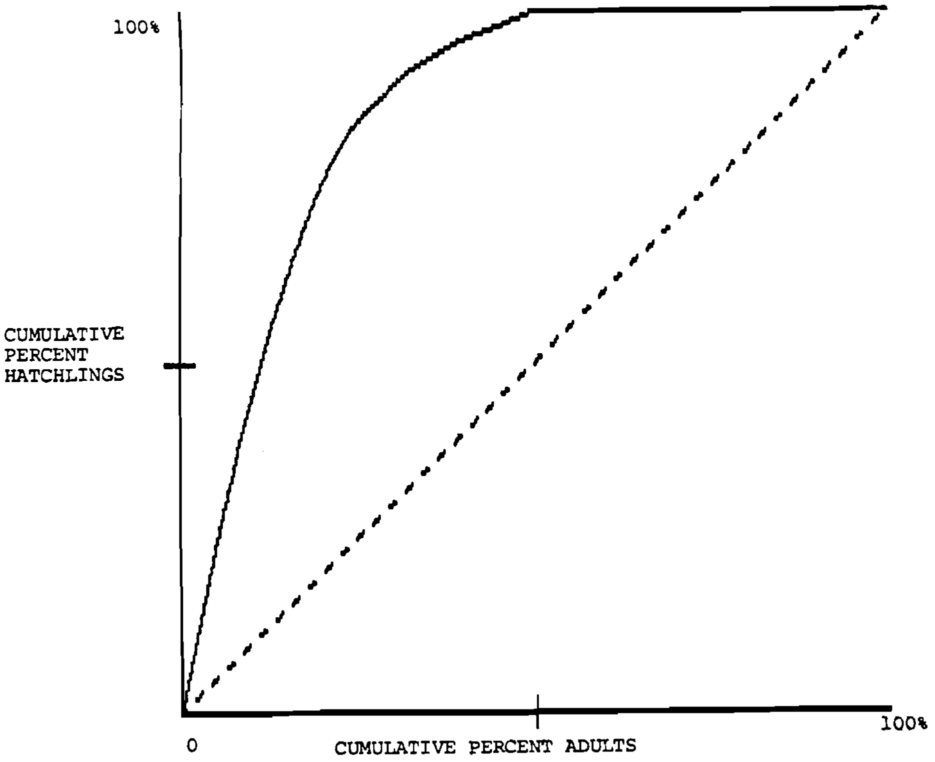


Figure 2a: Seasonal concentration of reproduction in male bullfrogs-1976. (Data from Howard, 1983)

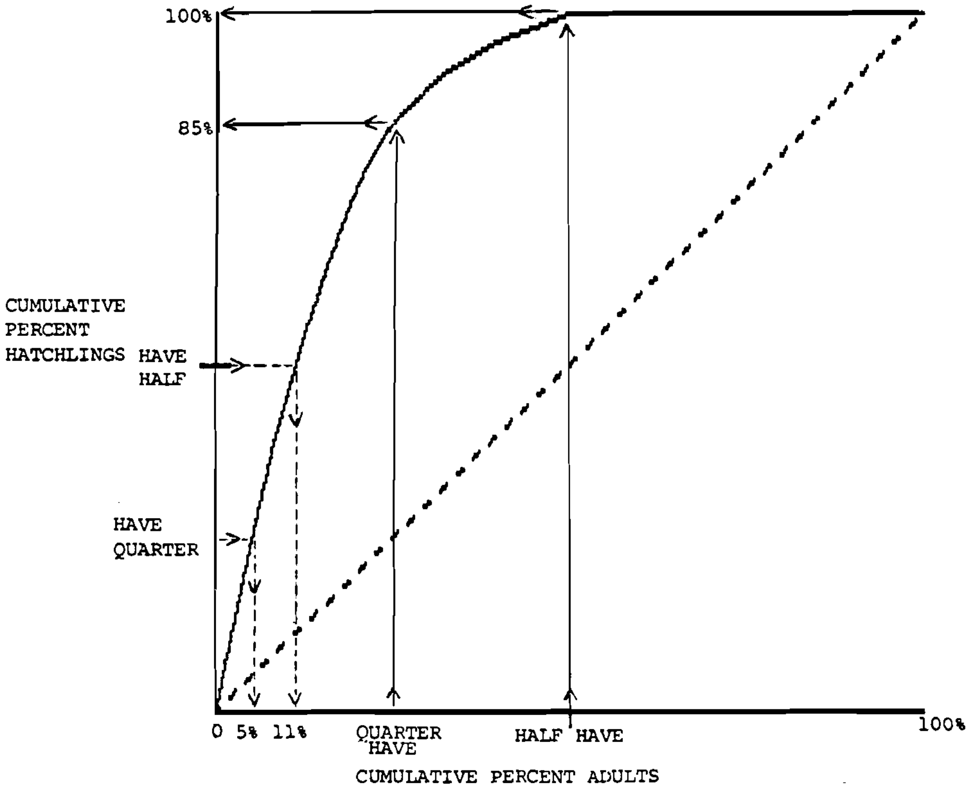


Figure 2b: Seasonal concentration of reproduction in male bullfrogs-1976. (Data from Howard, 1983)

Table 1. Calculations to derive a concentration curve from a frequency distribution.

Number of adults	Number of offspring per adult	Percent of adults	Cumulative percent of adults	Percent of offspring	Cumulative percent of offspring
1	29000	2.6	2.6	13.8	13.8
1	26000	2.6	5.3	12.4	26.2
1	23000	2.6	7.9	11.0	37.1
1	21000	2.6	10.5	10.0	47.1
1	19000	2.6	13.2	9.1	56.2
1	18000	2.6	15.8	8.6	64.8
1	15000	2.6	18.4	7.1	71.9
1	13000	2.6	21.2	6.2	78.1
1	10000	2.6	23.7	4.6	82.9
1	7000	2.6	26.3	3.3	86.2
2	5000	5.3	31.6	4.8	91.0
1	4000	2.6	34.2	1.9	92.9
3	3000	7.9	42.1	4.3	97.1
3	2000	7.9	50.0	2.9	100.0
19	0	50.0	100.0	0.0	
38	210000	100.0		100.0	

NOTE: **Apparent inconsistencies in addition are due to rounding. All calculations were done prior to rounding.**

Some Background on Concentration Curves and Have-Statistics

Economists use concentration curves to study the inequality in the distribution of income or wealth in a population and to study the concentration of business activity, e.g., what percentage of the output of the steel industry is produced by what percentage of the steel manufacturers? Such curves, which were proposed by Lorenz (1905), are often called Lorenz curves. In summarizing concentration curves, economists frequently use the Gini coefficient, which is equal to twice the area between the concentration curve and the 45 degree diagonal: the Gini coefficient varies from zero to one depending on how bowed the curve is.

Typically economists rank individuals on the horizontal axis from lowest income, wealth, or sales to greatest, so that the concentration curve falls below the 45 degree line and is cupped upward. In the life sciences, especially in studies where the most numerous, most prolific or most dominant individuals or species are of the greatest interest, it seems more natural to use a ranking from greatest to least. Margalef (1958) and Whittaker (1958) use this ranking for dominance-diversity curves and Patil and Taillie (1979a and b) use it in their "intrinsic diversity profiles"; Taillie (1979) draws a "Lorenz curve" for species diversity that follows this ranking. A ranking from greatest to least where individuals (or species or corporations) are on the x-axis and their cumulative importance is given on the y-axis leads to a concentration curve that lies above the diagonal and is cupped downward. By analogy to the terms "left-hand" and "right-hand" cumulative probability distributions, the concentration curve used by economists might be called a left-hand curve in contrast to the right-hand curve used in this paper.

Concentration curves have been employed by a variety of analysts studying different aspects of dominance and evenness. Geographers, for instance, use such curves to represent the spatial concentration of people in some land area: e.g., to analyze the proportion of a population occupying the most densely populated land. Political scientists use the curve to analyze such issues as racial imbalance in public schools and legislative malapportionment (Alker 1983) and in studies of influentials or elites "who get the most of what there is to get" (Lasswell 1958).

One of the earliest articles by an economist (Persons 1908) on the concentration of wealth suggests that economists may have been inspired by biologists: "The statistical problem before the economist in determining upon a measure of the inequality in the distribution of wealth is identical with that of the biologist in determining upon a measure of the inequality in the distribution of any physical characteristic". Thus, the concept of concentration may, to use Lewontin's (1984) phrase, "be stolen goods that really belonged to us in the first place". Nonetheless, it has been economists who have made the most use of concentration curves and related measures and biologists who have taken the least advantage of them.

Have-statistics are frequently used in summarizing concentration curves; they are sometimes referred to as fractiles or percentiles. An economist, for instance, might summarize the concentration of wealth by saying that the top ten percent of the population have seventy percent of the total wealth. In a recent news report in *Science* (Kolata 1985), the head of the U.S. National Cancer Institute is quoted as saying "50 percent of all cures through chemotherapy occur in 10 percent of all cancer patients". Some use of such statistics also occurs in biological analyses. For example: Le Boeuf and Peterson (1969, p. 91), in their discussion of reproductive success in male elephant seals, note that "Four percent of the males inseminated 85 percent of the females". Lill (1974, p.5), notes that in one of his studies of the white-bearded manakin, "a single male performed 73.3% of all observed copulations and 30% of the resident male population performed 95.4%".

Some standardization of such have-statistics would be helpful in comparing different studies. We have found two have-statistics especially informative in our research, the "havehalf" and the "havenone". The havehalf measure gives the percentage of the x's (e.g., individuals, females, males, species or, more generally, the "haves") that have half of y (e.g., total matings, hatchlings, surviving daughters, nesting sites, territorial area, biomass or, more generally, the "hads"); the havenone measure gives the percentage that have none. Two other measures that we have used on occasion are the havesome and the halfhave. The havesome, which is the complement of the havenone and which might just as well be called the haveall, gives the percentage of the x's that have at least some y (and hence collectively have all of y) and the halfhave gives the percentage of y that the top half of the x's have. Finally, we have also used the havequarter and the quarterhave. These measures may be particularly enlightening when there is extreme concentration. In general, the symbols h_y and ${}_xh$ might be used to denote any have-y and x-have statistics.

The have-y statistics are measures of the evenness of a distribution and the x-have statistics are measures of unevenness, i.e., dominance. Consider, for instance, the havehalf and the halfhave and assume that the distribution in question concerns variation in the number of offspring. If the distribution is perfectly even, so that all individuals have the same number of offspring, the havehalf and halfhave are both equal to 50%: half the individuals have half the offspring. If, on the other hand, the distribution is unequal, then the havehalf might be 10% (i.e., 10% of the individuals have half the offspring) and the halfhave might be 80% (i.e., half the individuals have 80% of the offspring).

Although a range of have-statistics undoubtedly have their uses, in most ecological analyses it may be sufficient, and it would certainly aid communication, to stress a standard set of measures, in particular the havehalf and the havenone. In biological studies, especially those concerned with the relationship between one generation and the next, the havehalf and havenone measures not only have a clear and natural significance, but are also readily comprehensible and intuitively meaningful. With some practice in using these two measures, an ecologist might find some additional have-statistics of some value: we have grown accustomed to using not only the havehalf and havenone, but also the havequarter, quarterhave and halfhave, and in the examples below we present all of these statistics.

Female Red Deer

A second example of the use of concentration curves and have-statistics can be developed from data on the estimated lifetime reproductive success of female red deer. These data, compiled by Clutton-Brock and his colleagues (1982), are based on their authoritative study of red deer on the Scottish island of Rhum. Data of this kind, which are difficult to gather and hence sparse, are invaluable in studies of evolution and natural selection because they incorporate individual differences in both reproduction and survival. Figure 3 presents Clutton-Brock's frequency distribution curve for the estimated lifetime reproductive success of a cohort of hinds measured by their number of offspring surviving to one year of age. The main factors determining the distribution are, according to the analysis of Clutton-Brock *et al.*, adult lifespan, fecundity, and calf mortality. Calf mortality depends on the mother's condition, matriarchal group size and population density as well as calf birth date and weight.

Figure 4 presents three concentration curves and some have-statistics that summarize these data. The most bowed curve includes all the data from the distribution curve and displays the estimated lifetime differential reproductive success of all of the females in the cohort. Most of the females without offspring are those which died before reproductive age. These animals largely comprise the flat top of the concentration curve. The less bowed curve presents the concentration of reproduction among those females surviving to reproductive age. The least bowed curved represents the concentration of reproduction among mothers only.

Note that among all of the female deer 23% had half the surviving offspring, among those surviving to reproductive age 35% percent had half the surviving offspring, and among the mothers 36% had half the offspring. The jump from 23% to 35% is due to the large proportion of the females, more than a third, that did not survive to reproductive age; almost all the deer that did survive to reproductive age became mothers, as reflected in the similarity of the 35% and 36% statistics and the two underlying concentration curves.

As this example illustrates, it may be of interest to examine two havehalf measures--among all the individuals as well as among only the individuals that had some offspring, matings, nesting sites, etc. The havenone measure provides a kind of bridge between these two havehalf measures. Indeed, there is a simple mathematical relationship between the two havehalf measures and the havenone measure. Suppose the data concern females, mothers, and children. If a proportion p of the females are mothers (i.e., have some children) and a proportion q of

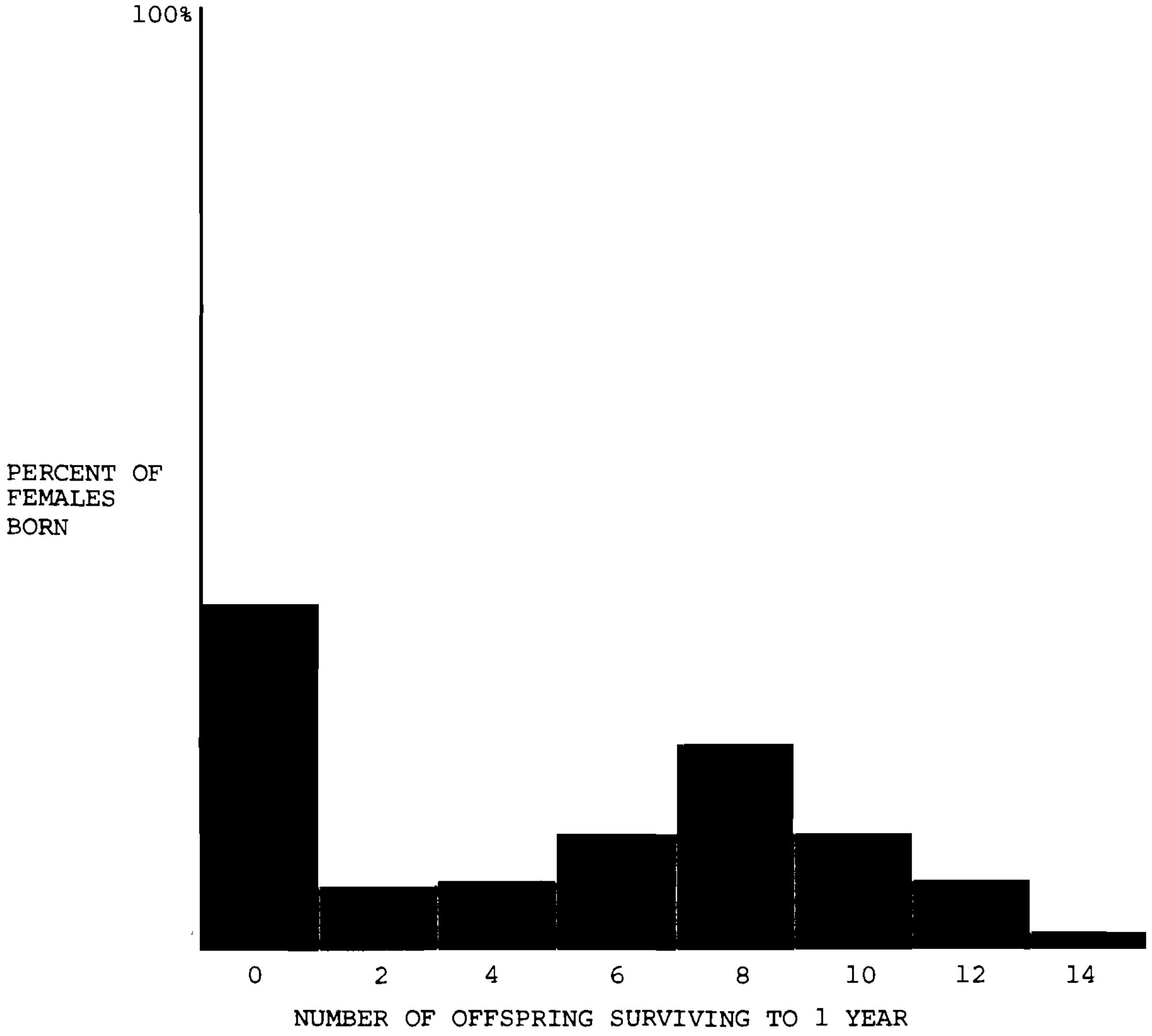
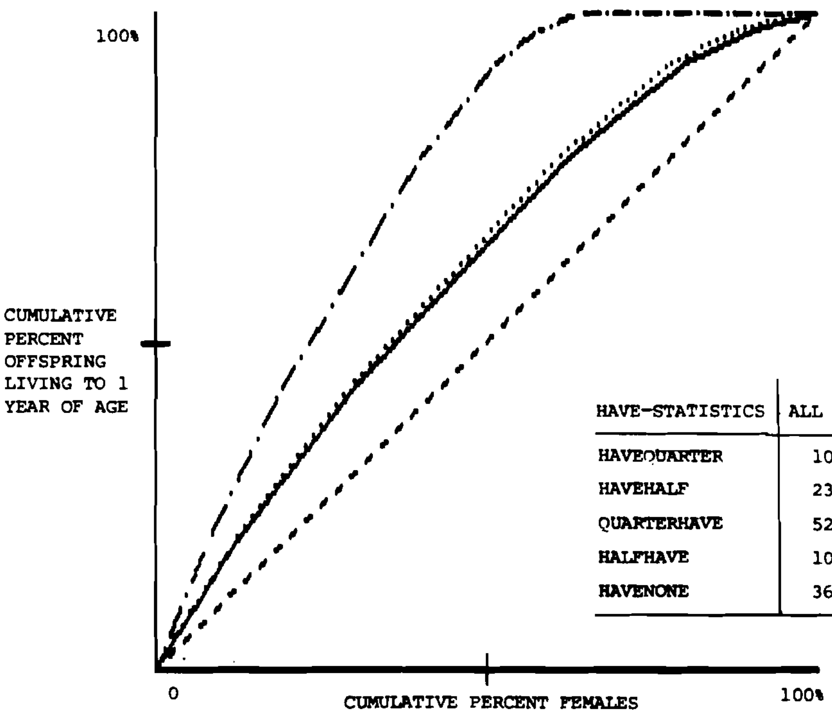


Figure 3: Estimated distribution of lifetime reproductive success in red deer hinds.
(Redrawn from Clutton-Brock et al, 1982)



HAVE-STATISTICS	ALL FEMALES	FEMALES SURVIVING TO REPRODUCTIVE AGE	MOTHERS
HAVEQUARTER	10.4%	15.6%	16.1%
HAVEHALF	23.3%	35.1%	36.2%
QUARTERHAVE	52.9%	37.8%	36.8%
HALFHAVE	10.3%	66.3%	64.6%
HAVENONE	36.0%	3.0%	0.0%

Figure 4: Estimated concentration of lifetime reproductive success in red deer hinds for females:
- Born -·-·-·
- Surviving to reproductive age ······
- Who become mothers ————

(Data from Clutton-Brock et al, 1982)

the mothers have half the children, then clearly the proportion p times q of the females have half the children. In other words, if the havesome is p and the mothers' havehalf is q, then the females' havehalf is pq. Since the havenone is just one minus the havesome, this expression could also be expressed as follows: if the havenone is r and the mother's havehalf is q, then the females' havehalf is $q(1-r)$.

Comparative Analysis: Bullfrogs Over Time

Concentration curves and have-statistics are especially useful in comparing related sets of data concerning some aspect of diversity or evenness. An instructive example concerns the differences over time in the reproductive success of male bullfrogs. Howard (1983) gives frequency distributions of the estimated reproductive success of male bullfrogs, as measured by number of hatchlings produced, in three successive years from 1976 to 1978; these distributions are shown in Figures 5a, b and c. Re-expressing these distributions as concentration curves, as in Figure 6, can help reveal the differences and similarities among the three years. The most striking similarity is that in all three years roughly 50% of the frogs had no hatchlings. This is immediately apparent in the concentration curves. Beyond this, the curves for 1976 and 1978 are rather similar and indicate a higher degree of concentration of reproductive success than in 1977.

Note, however, that among the prolific frogs accounting for half the offspring, the 1976 and 1978 curves are relatively dissimilar, and almost as far from each other as the 1976 curve is from the 1977 curve. Indeed, the havehalf statistics for 1978, 1976, and 1977, respectively, are 8.4%, 11.4%, and 16.2%. Thus, reproduction in 1978 was dominated by half as many individuals as in 1977, with 1976 falling roughly halfway in between. The higher degree of concentration of reproduction in 1976 compared with 1977 is not readily apparent on the frequency distributions in Figures 5a and b. On the other hand, the importance of the top two reproducers in 1978 is clearly indicated in the frequency curve in Figure 5c, although an analyst not concerned with or alert to the reproductive significance of these two frogs might unconsciously dismiss them as outliers.

Many frequency distributions in biological studies are similar to the frog distribution in being highly skewed: most of the frogs have few if any offspring, a few frogs are prolific. Concentration curves highlight the importance of the few, sometimes overlooked frogs in the tail-end of a frequency distribution by, in effect, weighting each frog by its number of offspring. In many biological studies, especially those concerned with natural selection or evolution, what is of prime interest is the small group of individuals who are producing most of the offspring, who control most of the nesting sites, who have most of the biomass, etc. Concentration curves and have-statistics are a natural way of directly focusing on such concerns.

Additional Examples of Comparative Analysis

To further illustrate the use of concentration curves and have-statistics for comparative analysis, we briefly describe in this section a potpouri of six suggestive examples.

1. *Mating vs. Zygote vs. Hatchling Success for Male Bullfrogs*

Figure 7, which is based on data given in an earlier study by Howard (1979), displays a set of curves for the concentration of estimated reproductive success during three successive stages of reproduction for a mixed cohort of male

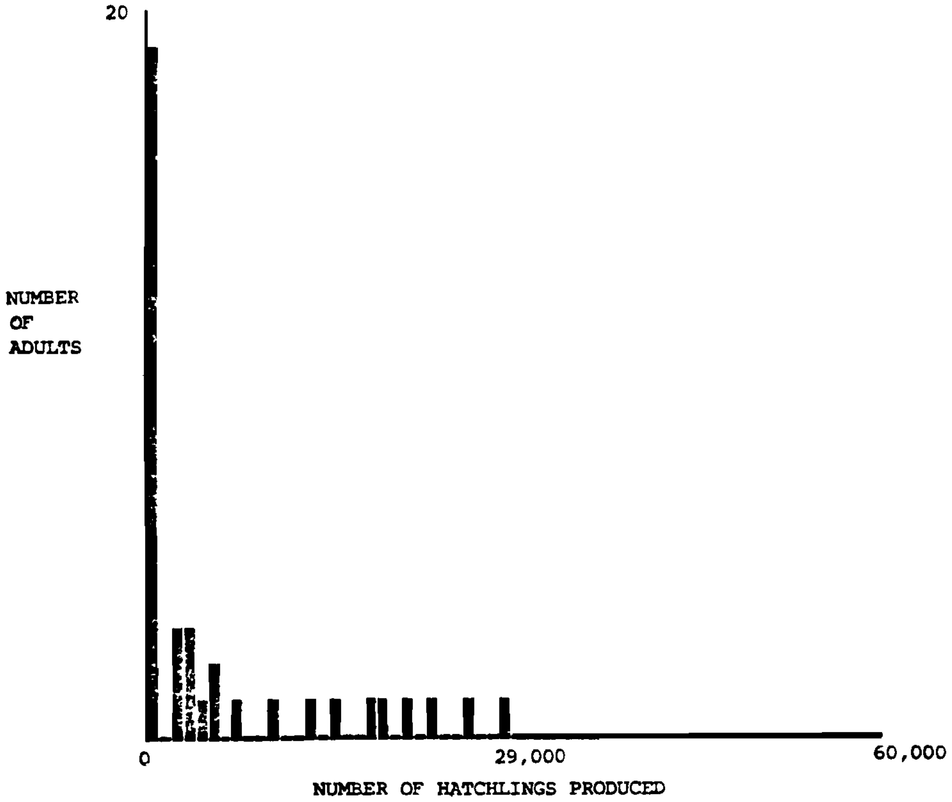


Figure 5a: Estimated seasonal reproductive success in male bullfrogs-1976. (Redrawn from Howard, 1983)

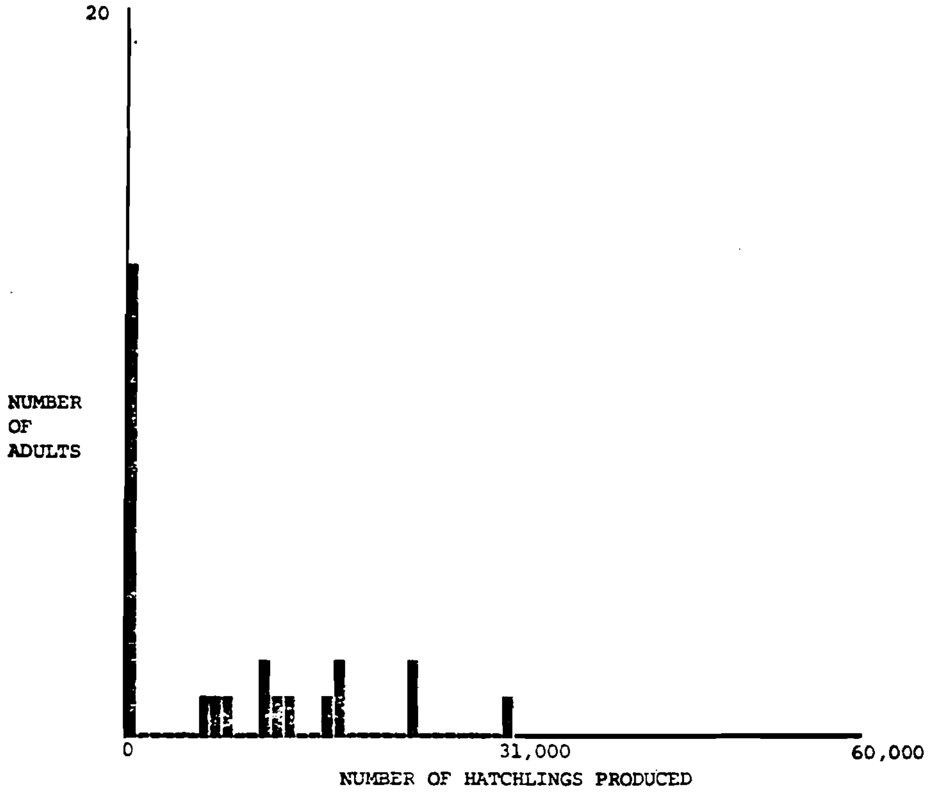


Figure 5b: Estimated seasonal reproductive success in male bullfrogs-1977. (Redrawn from Howard, 1983)

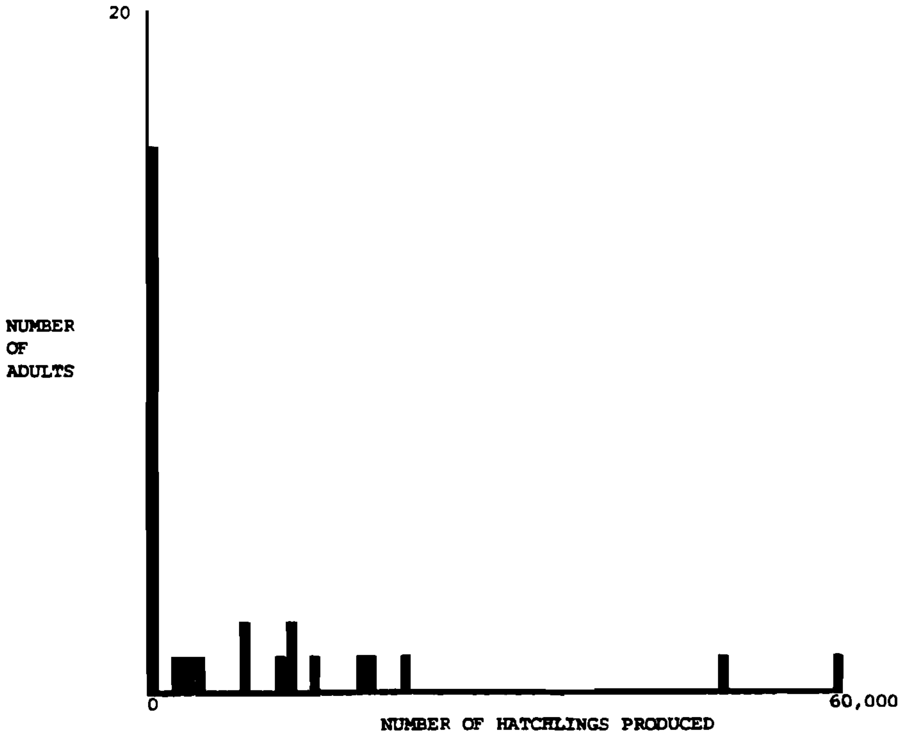
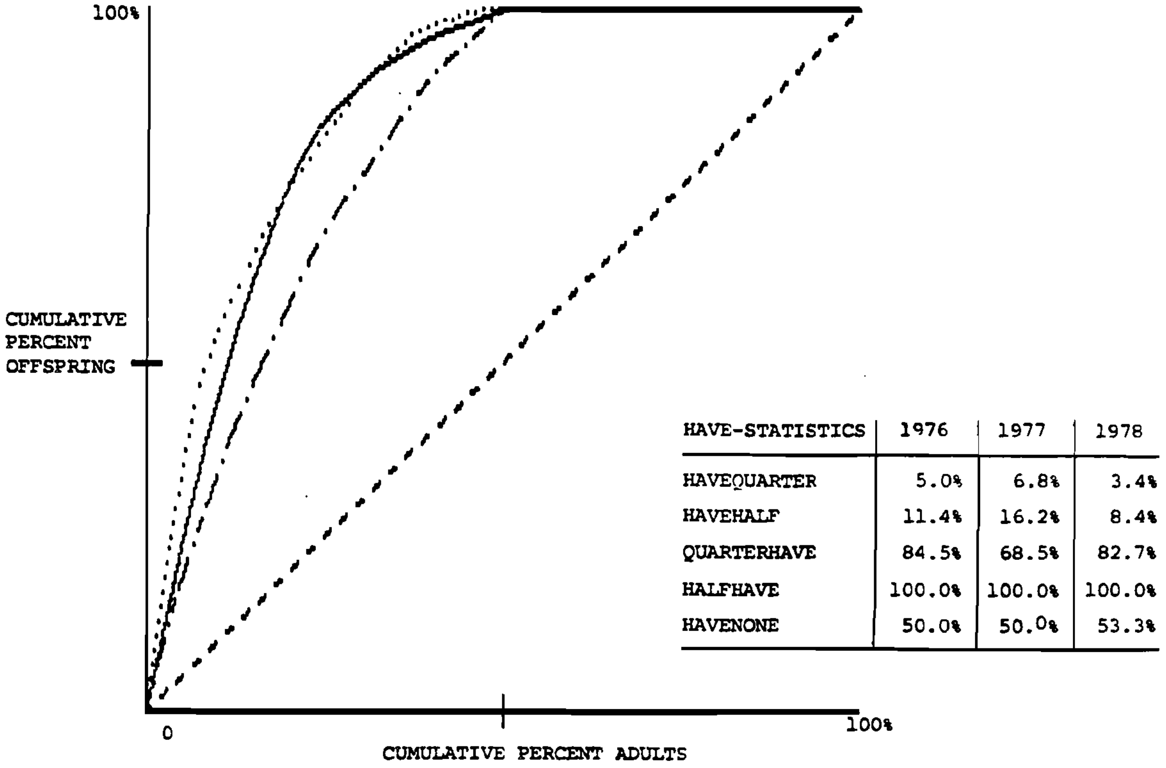


Figure 5c: Estimated seasonal reproductive success in male bullfrogs-1978. (Redrawn from Howard, 1983)

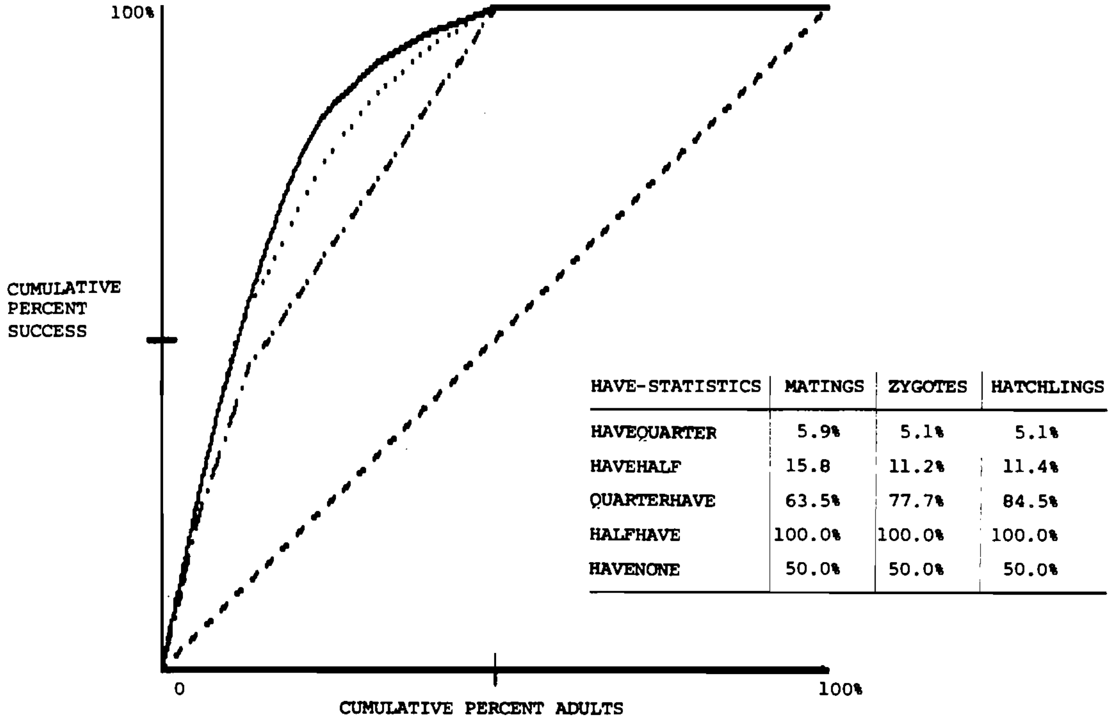


HAVE-STATISTICS	1976	1977	1978
HAVEQUARTER	5.0%	6.8%	3.4%
HAVEHALF	11.4%	16.2%	8.4%
QUARTERHAVE	84.5%	68.5%	82.7%
HALFHAVE	100.0%	100.0%	100.0%
HAVENONE	50.0%	50.0%	53.3%

Figure 6: Seasonal concentration of estimated reproductive success in male bullfrogs:
- 1976 ————
- 1977 —·—·—·—
- 1978 ·······

(Data from Howard, 1983)

bullfrogs. The curves are increasingly bowed. Thus in this population, reproductive success becomes increasingly concentrated in fewer individuals at stages progressing from mating to zygote to hatchling success. Have-statistics for these three stages of reproductive success are given on the right of the figure. Note the change in the quarterhave measure: the top quarter have about five-eighths of the matings, about three-quarters of the zygotes, and close to seven-eighths of the hatchlings.

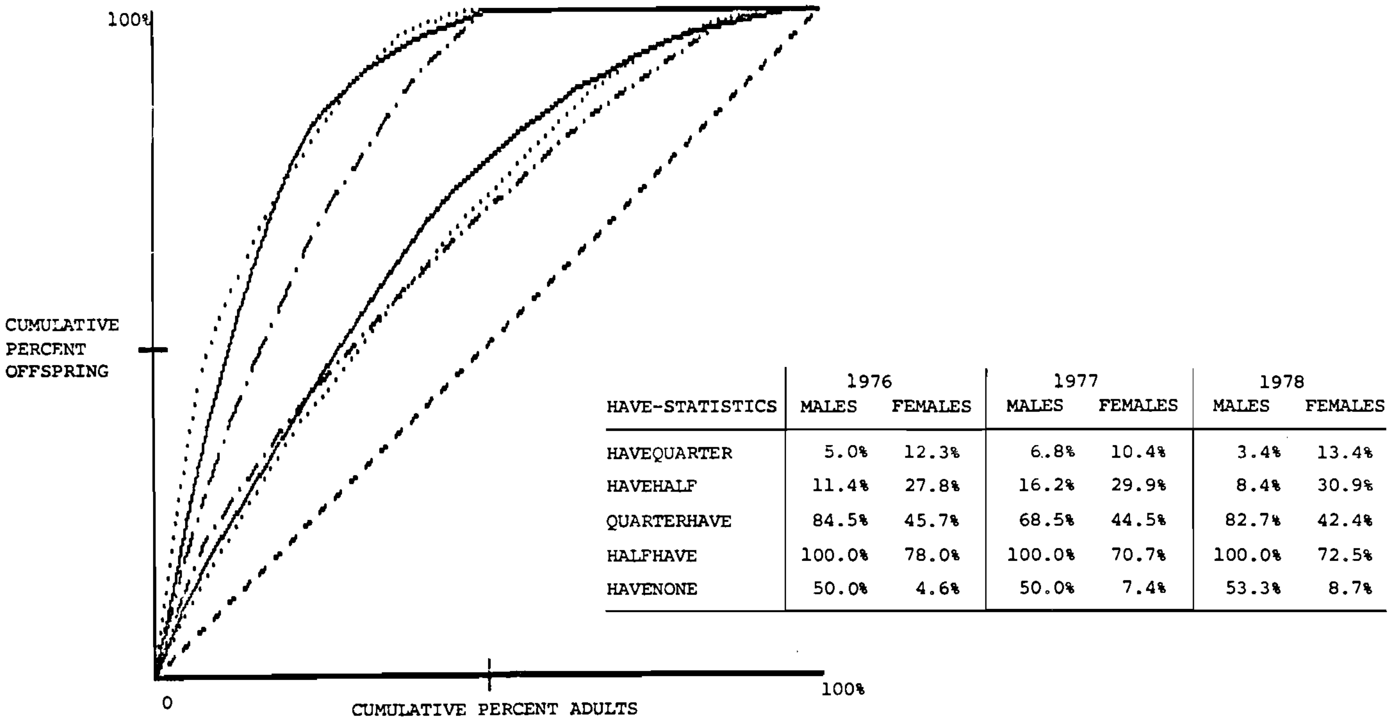


HAVE-STATISTICS	MATINGS	ZYGOTES	HATCHLINGS
HAVEQUARTER	5.9%	5.1%	5.1%
HAVEHALF	15.8	11.2%	11.4%
QUARTERHAVE	63.5%	77.7%	84.5%
HALFHAVE	100.0%	100.0%	100.0%
HAVENONE	50.0%	50.0%	50.0%

Figure 7: **Seasonal concentration of estimated reproductive success in male bullfrogs for advancing stages of reproduction:**
- mating —·—·—·—
- zygote ·······
- hatchling ————
(Data from Howard, 1979)

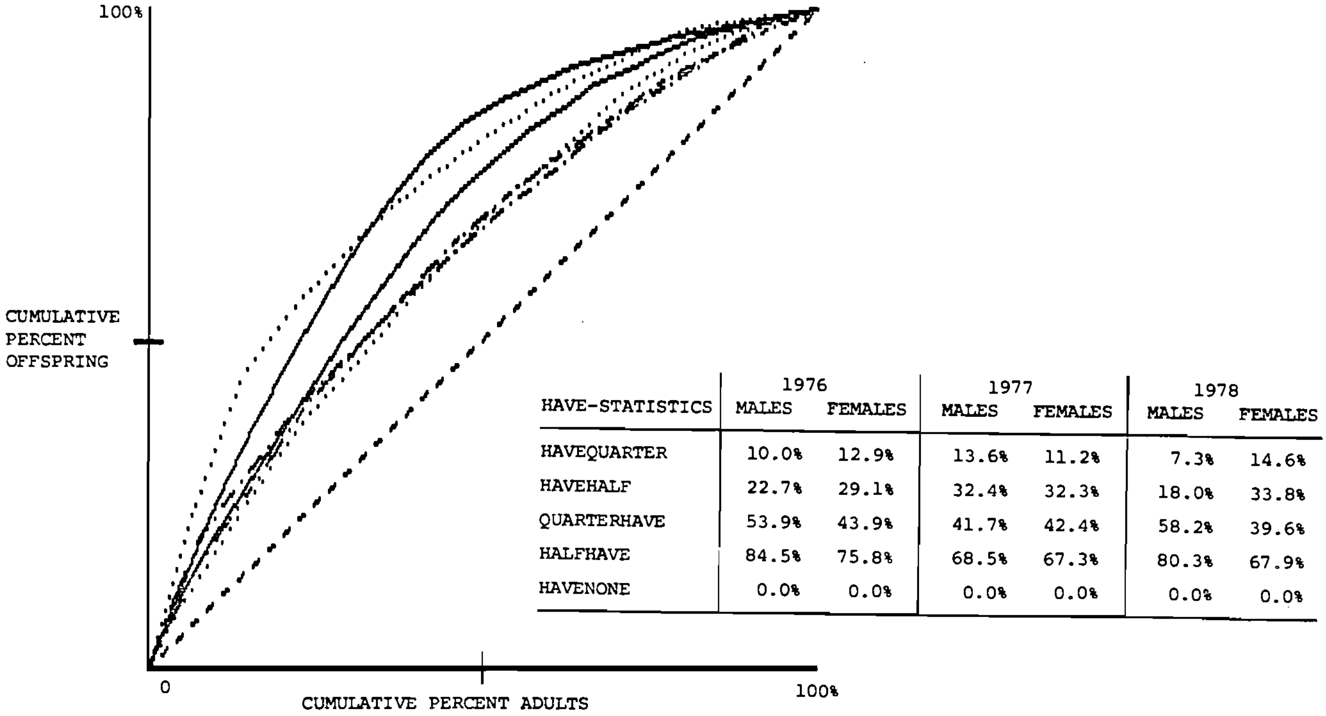
2. *Female vs. Male Bullfrogs*

Figure 8a, which is based on Howard's (1983) data, illustrates the concentration of seasonal reproductive success for female vs. male bullfrogs, for three successive years. The figure shows the greater concentration of male versus female seasonal reproductive success for this sample, mixed cohort, population. Figure 8b is drawn to include only those male and female animals contributing to reproduction--thus excluding half of the males and one or two females each year.



HAVE-STATISTICS	1976 MALES	 FEMALES	1977 MALES	 FEMALES	1978 MALES	 FEMALES
HAVEQUARTER	5.0%	12.3%	6.8%	10.4%	3.4%	13.4%
HAVEHALF	11.4%	27.8%	16.2%	29.9%	8.4%	30.9%
QUARTERHAVE	84.5%	45.7%	68.5%	44.5%	82.7%	42.4%
HALFHAVE	100.0%	78.0%	100.0%	70.7%	100.0%	72.5%
HAVENONE	50.0%	4.6%	50.0%	7.4%	53.3%	8.7%

Figure 8a: Seasonal concentration of estimated reproductive success in male (3 outer curves) and female bullfrogs:
- 1976 ————
- 1977 —·—·—·—
- 1978 ·······
(Data from Howard, 1983)



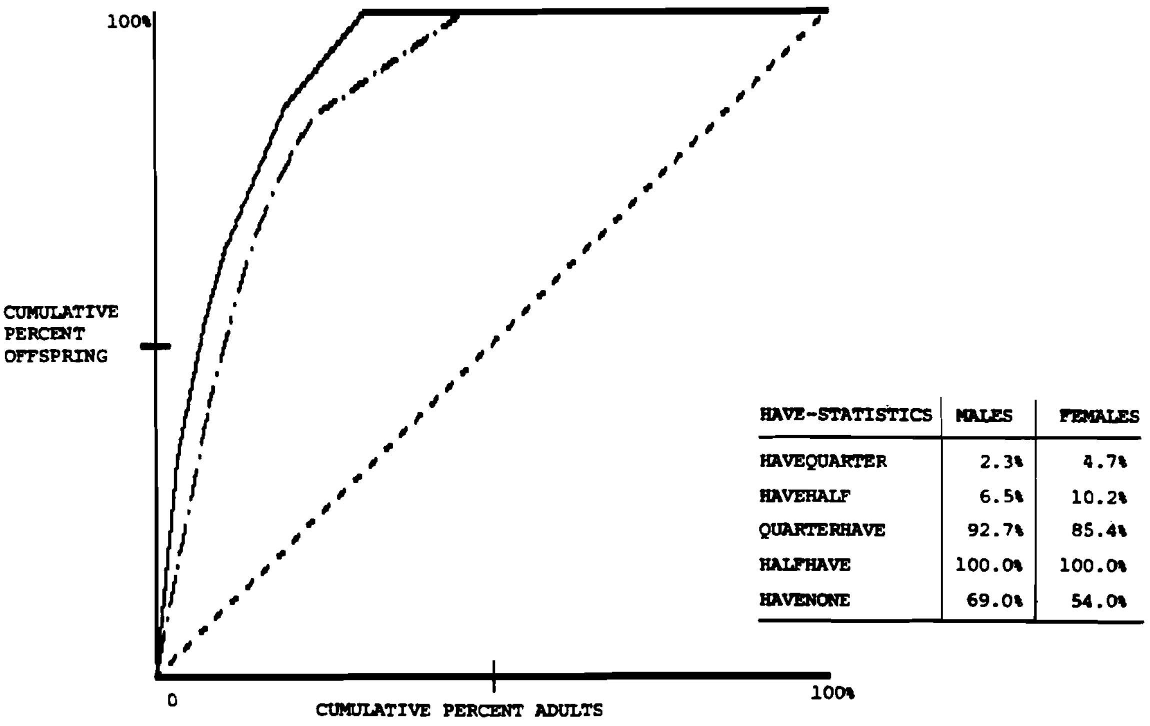
HAVE-STATISTICS	1976 MALES	 FEMALES	1977 MALES	 FEMALES	1978 MALES	 FEMALES
HAVEQUARTER	10.0%	12.9%	13.6%	11.2%	7.3%	14.6%
HAVEHALF	22.7%	29.1%	32.4%	32.3%	18.0%	33.8%
QUARTERHAVE	53.9%	43.9%	41.7%	42.4%	58.2%	39.6%
HALFHAVE	84.5%	75.8%	68.5%	67.3%	80.3%	67.9%
HAVENONE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 8b: Seasonal concentration of estimated reproductive success for male and female bullfrogs having some offspring:
- 1976 ————
- 1977 —·—·—·—
- 1978 ·······
(Note: Male curves lie outside of female curves each year except 1977 where they overlap)
(Data from Howard, 1983)

When the havenone component is eliminated, the differential between males and females is substantially reduced. In fact, the curves for 1977 for males and females are hardly distinguishable. The curve for 1976 females shows more differential reproductive success, now more bowed than the 1977 male curve. The curves for 1976 and 1978 males still show the greatest degree of differential success.

3. *Predicted Lifetime Reproductive Success of Male and Female Bullfrogs*

Howard (1983) used his data on estimated seasonal reproductive success in bullfrog populations to construct a model of their predicted lifetime reproductive success. This is particularly interesting as it removes effects due to mixed cohorts in the seasonal data. Concentration curves and have-statistics based on his results are displayed in Figure 9. The data include all males and females who survived to age one. The concentration curves clearly display what Howard concludes, namely that there is relatively little intersexual differential in lifetime reproductive success. This is a surprising result in light of the substantial differences between male and female seasonal reproductive success, as shown in Figure 8a.



HAVE-STATISTICS	MALES	FEMALES
HAVEQUARTER	2.3%	4.7%
HAVEHALF	6.5%	10.2%
QUARTERHAVE	92.7%	85.4%
HALFHAVE	100.0%	100.0%
HAVENONE	69.0%	54.0%

Figure 9: Predicted concentration of lifetime reproductive success for bullfrogs:
- males ————
- females —·—·—
(Data from Howard, 1983)

The reason for the discrepancy can be explained as follows. The high degree of concentration among females is due to the large proportion of females, some 54%, with no offspring: many females are not sexually mature at age one and die before reproductive age. In Howard's seasonal data only reproductively mature animals are included, so that these immature females are excluded. Virtually all the females that are mature produce at least some young during a breeding season. Hence, the seasonal data show much less concentration of reproductive success among females than do the lifetime data.

On the other hand, all one year old males are mature and are included in Howard's seasonal data. Unlike females, however, many males fail to produce young during their first breeding season and then die before they have another breeding opportunity. Indeed, males and females follow different life history trajectories: females accumulate reproductive success during each mature year of life whereas males achieve most of their success in the latter years of life when they have grown to a large size. The lifetime reproductive success of the most prolific male tends to roughly equal that of the most prolific female. As shown by the concentration curves, the result is that the concentration of predicted lifetime reproductive success among males is only moderately greater than the concentration among females.

4. *Female vs. Male Fruit Flies*

Bateman (1948), in his classic paper on the distribution of mating success in Drosophila, hypothesized that female reproductive success is limited by egg production and male reproductive success by the number of eggs they fertilize. He suggested that matings will be much more variable among males than among females. Whereas among the males, some individuals will have many matings and many will have none, among the females, mating success will be more evenly distributed. In effect this situation can be viewed as a competition for a limited resource and the degree of variability in male mating success indicative of the degree of competition. Concentration curves are an appropriate way of summarizing the degree of this competition among males and the difference in variability of mating success for males versus females.

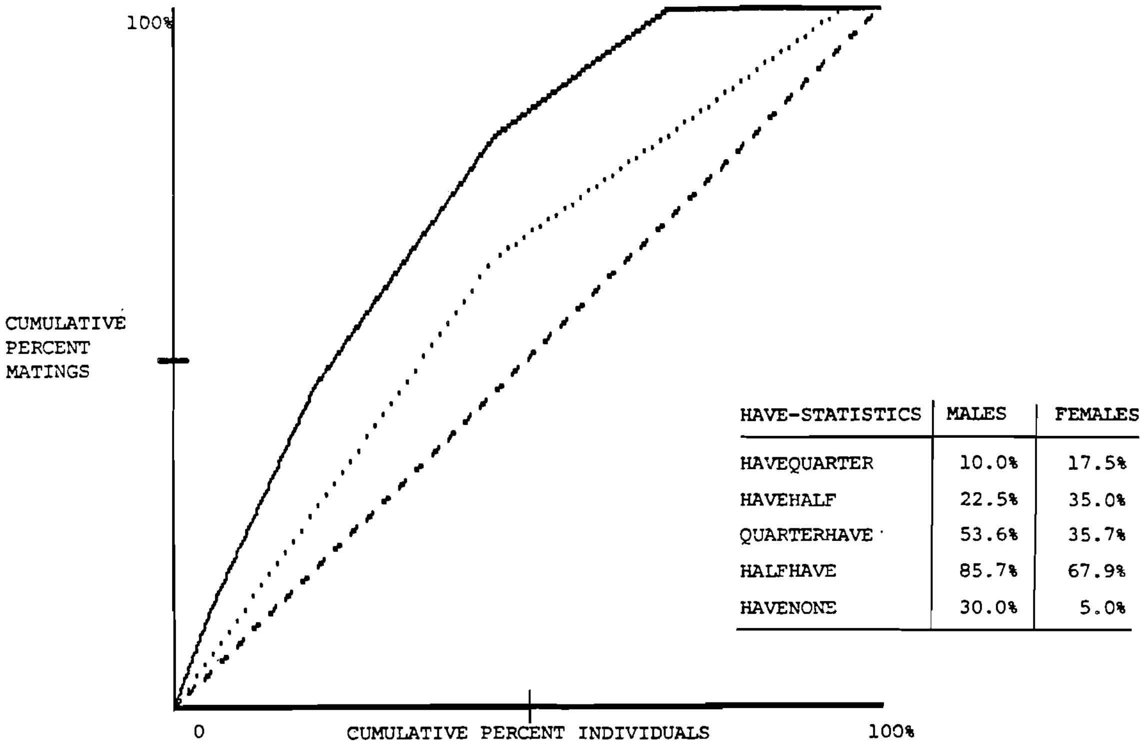
Figure 10a displays the concentration of reproductive success for male and female Drosophila, based on Bateman's data. Concentration of reproductive success is clearly higher among males, supporting Bateman's argument. Figure 10b shows the concentrations when males and females with no matings are removed. The curves are fairly close, indicating that most of the difference between the sexes is attributable to the large proportion of males, some 30 percent, with no matings compared with only 5 percent of the females.

5. *Male Sculpins on Natural vs. Experimental Plots*

Figure 11 illustrates the differential reproductive success of male sculpins on natural and experimental plots. Downhower and Brown (1980) suggest that, in this population, female choice and hence male reproductive success is more influenced by male than territorial quality. To test this hypothesis, the authors compared reproductive success on natural and experimental plots over a period of three years. During the first two years the experimental plots were heterogeneous and in the final year the experimental plot was homogeneous. The set of four graphs in Figure 11 clearly displays the similarity in reproductive success independent of territory quality. As shown in the fourth graph, the curves are all similar, supporting the Downhower hypothesis. Nonetheless, the first three graphs indicate that in the first two years, when the experimental plots were heterogeneous, the concentration of reproduction for the experimental plots was higher than for the natural plots, whereas in the final year, when the experimental plot was homogeneous, the concentration of reproduction for the experimental plots was lower than for the natural plots. This reversal possibly indicates some influence of territory quality.

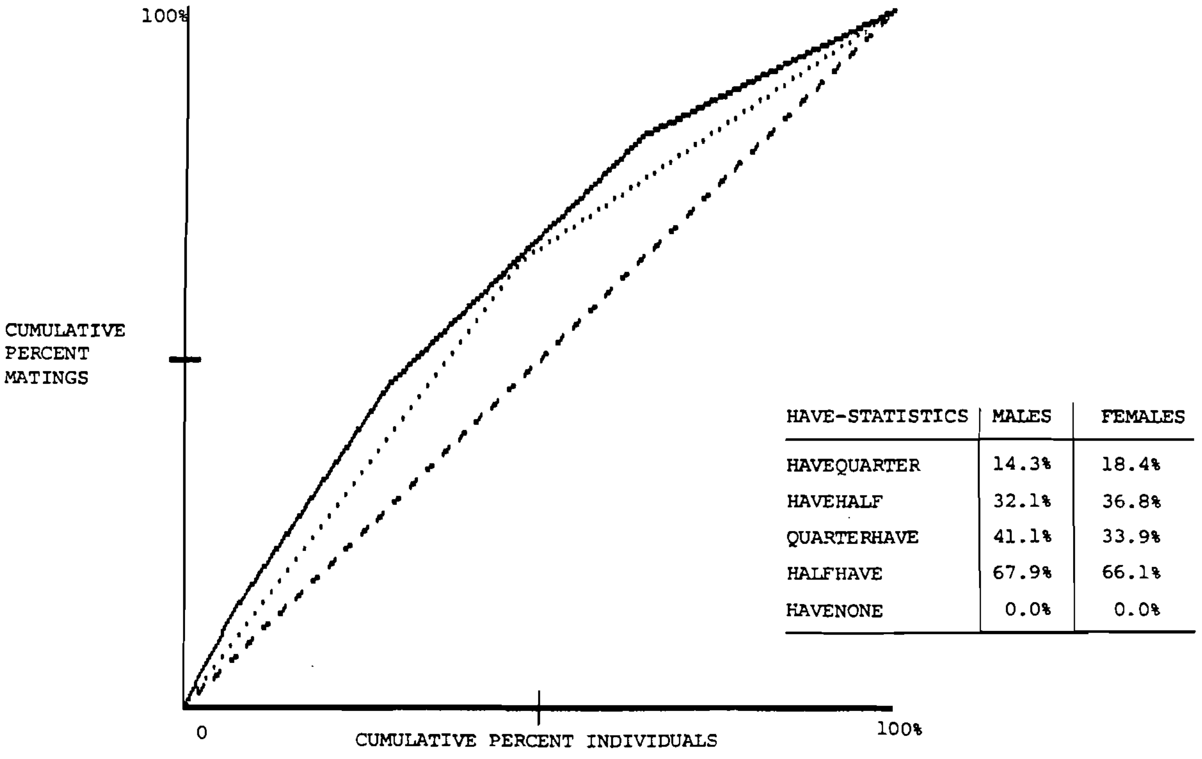
6. *Male Elephant Seals in an Expanding Population*

Figure 12 uses data from Le Boeuf's (1974) study of the distribution of reproductive success among male elephant seals in a population off the coast of California. The data for each year pertain to the ten top-ranking seals. The concentration curves and have-statistics clearly summarize the steady decline in concentration of male mating success over the 6 years of the study. This is a recently established colony and the population expanded over the course of the study. Le Boeuf hypothesizes that as the population stabilizes fewer males will again come to monopolize reproductive success.



HAVE-STATISTICS	MALES	FEMALES
HAVEQUARTER	10.0%	17.5%
HAVEHALF	22.5%	35.0%
QUARTERHAVE	53.6%	35.7%
HALFHAVE	85.7%	67.9%
HAVENONE	30.0%	5.0%

Figure 10a: Concentration of mating success for drosophila:
- males ————
- females ·······
(Data from Bateman, 1948)



HAVE-STATISTICS	MALES	FEMALES
HAVEQUARTER	14.3%	18.4%
HAVEHALF	32.1%	36.8%
QUARTERHAVE	41.1%	33.9%
HALFHAVE	67.9%	66.1%
HAVENONE	0.0%	0.0%

Figure 10b: Concentration of mating success for drosophila having some matings:
- males ————
- females ·······
(Data from Bateman, 1948)

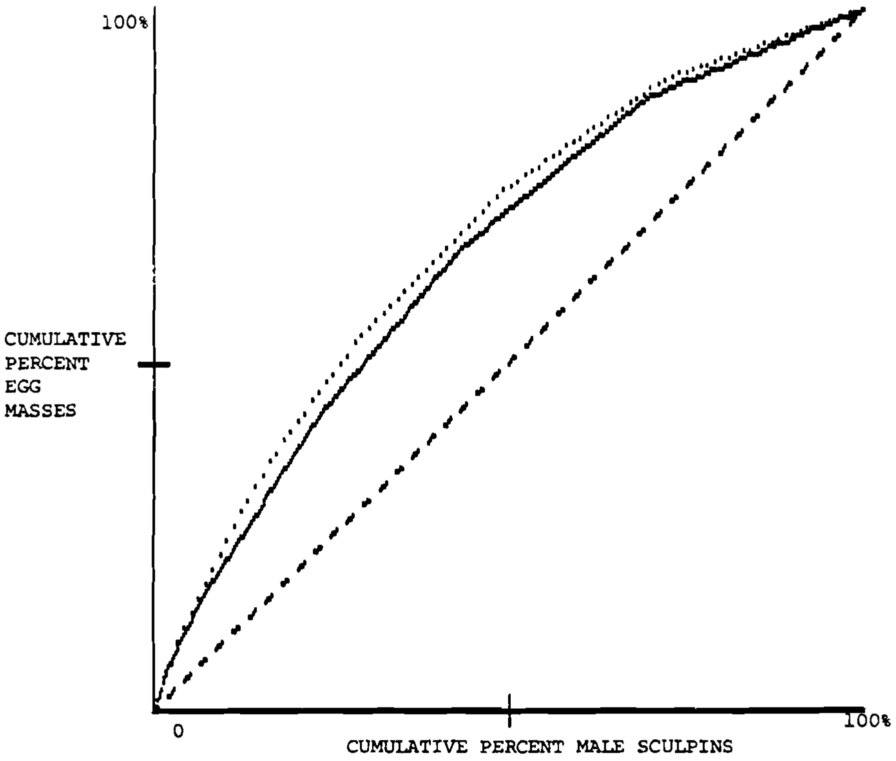


Figure 11a: Concentration of reproduction in male sculpins-1975:
- Natural plots ———
- Experimental plots ....
(Data from Downhower et al, 1980)

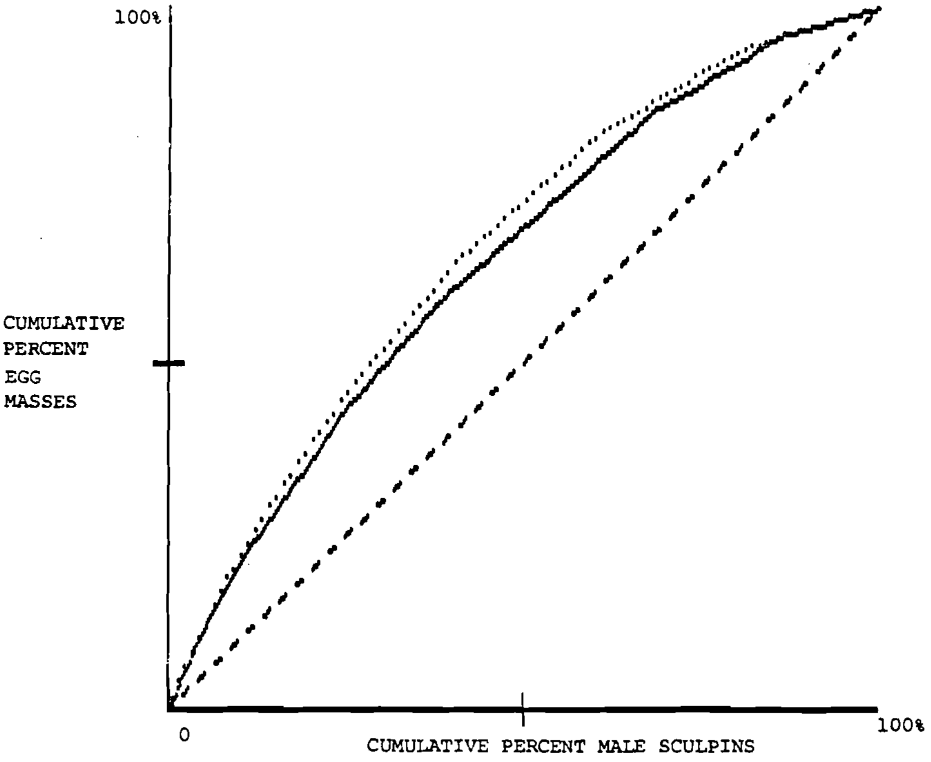


Figure 11b: Concentration of reproduction in male sculpins-1976:
- Natural plots ———
- Experimental plots ....
(Data from Downhower et al, 1980)

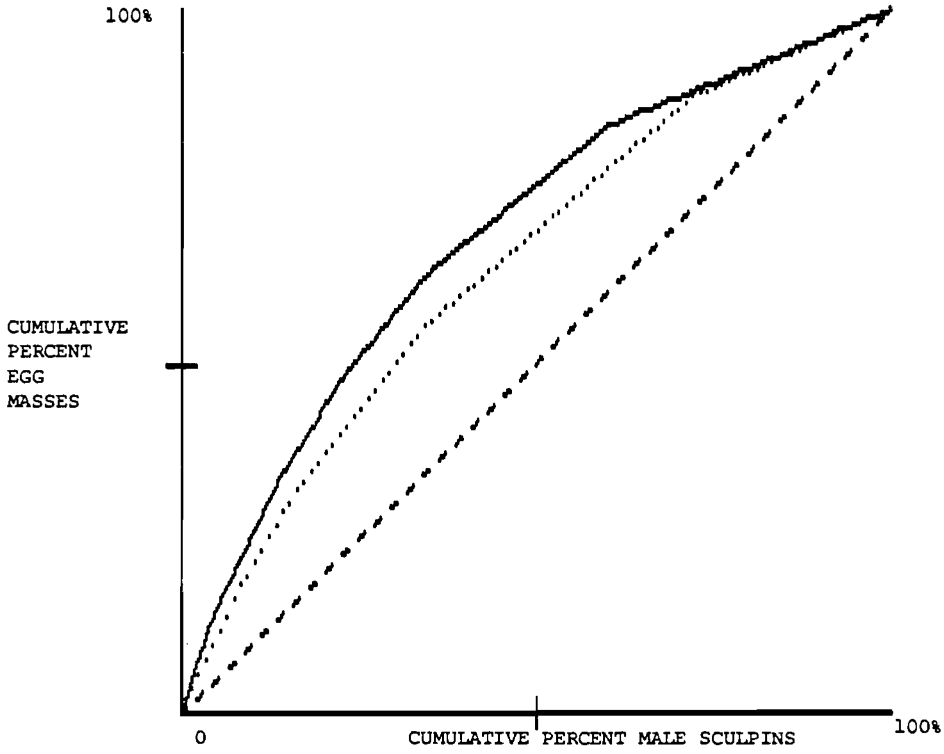
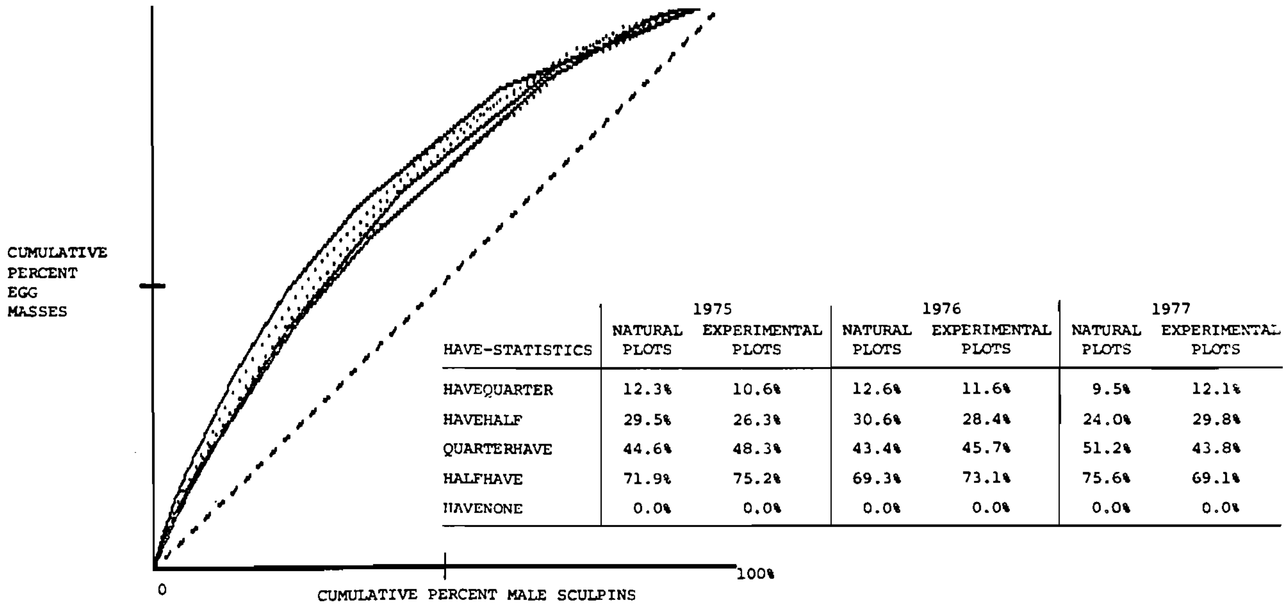
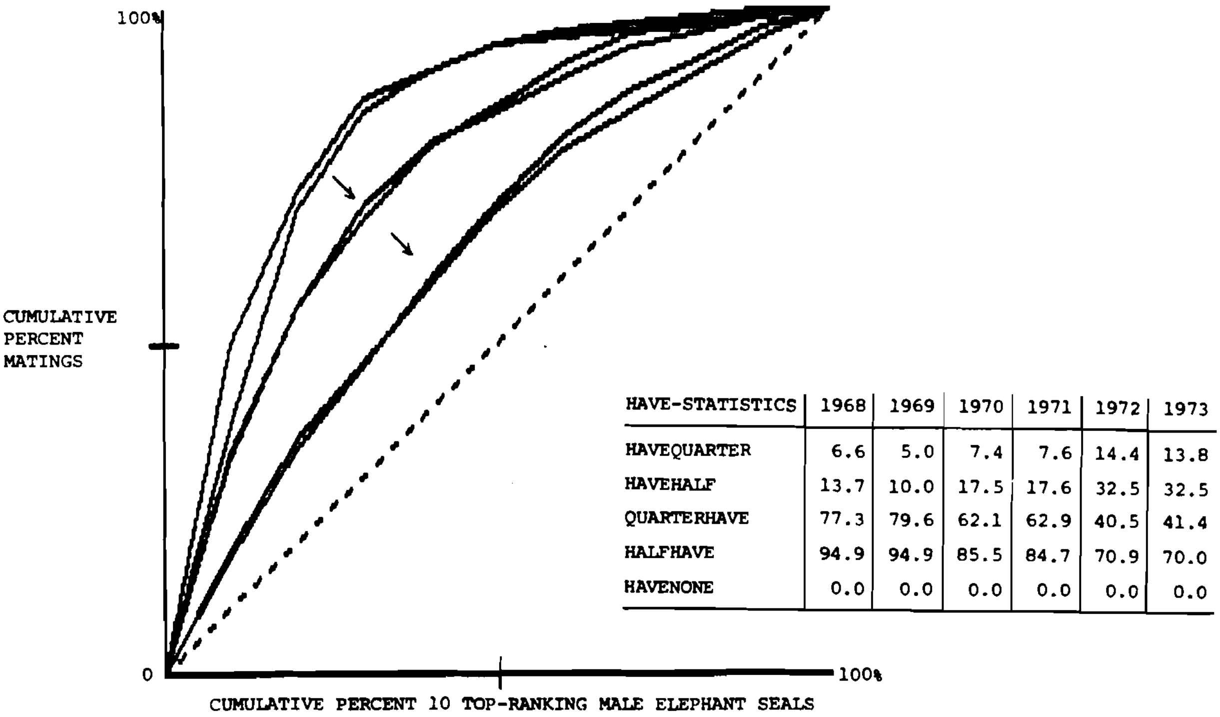


Figure 11c: Concentration of reproduction in male sculpins-1977:
- Natural plots ————
- Experimental plots ·······
(Data from Downhower et al, 1980)



	1975		1976		1977	
HAVE-STATISTICS	NATURAL PLOTS	EXPERIMENTAL PLOTS	NATURAL PLOTS	EXPERIMENTAL PLOTS	NATURAL PLOTS	EXPERIMENTAL PLOTS
HAVEQUARTER	12.3%	10.6%	12.6%	11.6%	9.5%	12.1%
HAVEHALF	29.5%	26.3%	30.6%	28.4%	24.0%	29.8%
QUARTERHAVE	44.6%	48.3%	43.4%	45.7%	51.2%	43.8%
HALFHAVE	71.9%	75.2%	69.3%	73.1%	75.6%	69.1%
HAVENONE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 11d: Concentration of reproduction in male sculpins-1975, 1976 & 1977:
- Natural plots ————
- Experimental plots ·······
(Data from Downhower et al, 1980)



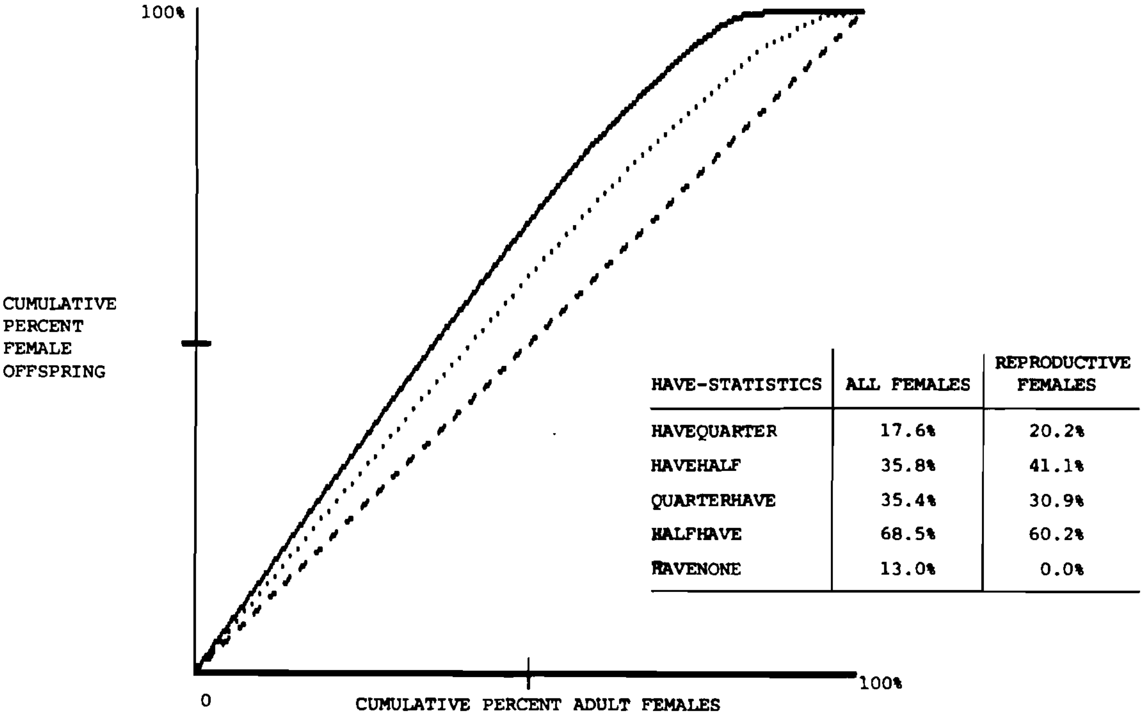
HAVE-STATISTICS	1968	1969	1970	1971	1972	1973
HAVEQUARTER	6.6	5.0	7.4	7.6	14.4	13.8
HAVEHALF	13.7	10.0	17.5	17.6	32.5	32.5
QUARTERHAVE	77.3	79.6	62.1	62.9	40.5	41.4
HALFHAVE	94.9	94.9	85.5	84.7	70.9	70.0
HAVENONE	0.0	0.0	0.0	0.0	0.0	0.0

Figure 12: Concentration of matings in male elephant seals over six years in an expanding population. (Note: The curve flattens as it moves from 68/69→ 70/71→ 72/73.) (Data from LeBoef, 1974)

Concentration Curves from Lifetables: Rice Weevils

Curves showing the concentration of reproduction can be readily calculated from lifetables that give a schedule of mortality and fertility, providing a convenient summary of lifetable data from a biologically relevant perspective. To illustrate this, Table 2, which presents mortality and fertility data for rice weevils, was used to produce Figure 13; the mortality and fertility rates in the second and fourth columns of the table were taken from Odum (1983, p.305). This figure indicates that under the optimal conditions of the study and assuming that all females' survival and reproduction chances would be governed by the lifetable values, there would be a fairly low concentration of reproduction: 36 percent of the females would have half the offspring and only 13 percent would have no offspring. If the 13 percent with no offspring are eliminated, the concentration curve among mothers falls remarkably close to the diagonal line of complete evenness: 41 percent of the mothers would have half the offspring (and the other 59 percent would have the other half).

To produce a concentration curve from lifetable data, the first step is to calculate a frequency distribution that gives the proportions of females with various numbers of offspring. Then the methods described above, in connection with Table 1, can be used to compute the concentration curve. Table 2 illustrates how the frequency distribution can be calculated. Column 3 of the table gives the proportion of females dying in each age category. Column 5 gives the cumulative number of (female) offspring for females who survive to the beginning of each age category. Under the assumption that females die halfway through an age category, the total number of offspring for a female that dies in an age category is given by the cumulative number up through the start of that age category plus half the value for the age category: Column 6 gives these numbers of offspring. For the last age category of 18.5 weeks or more, we assumed that females produced the full number of offspring, namely, one. Together, columns 3 and 6 provide the frequency distribution information needed to compute a concentration curve (using the method explained in Table 1). Further information about lifetable calculations can be found in Keyfitz (1968).



HAVE-STATISTICS	ALL FEMALES	REPRODUCTIVE FEMALES
HAVEQUARTER	17.6%	20.2%
HAVEHALF	35.8%	41.1%
QUARTERHAVE	35.4%	30.9%
HALFHAVE	68.5%	60.2%
HAVENONE	13.0%	0.0%

Figure 13: Concentration of reproductive success for a laboratory population of rice weevils studied under optimum conditions:
- all females ————
- females having some offspring ·······
(Data from Odum, 1983; after Birch, 1948)

It is crucial to note that the calculations in Table 2 are based on the assumption that all individuals, at birth, will experience, over the course of their lives, the mortality and fertility rates given in the lifetable. This assumption implies that all individuals have equal Darwinian fitness, in the sense that each individual, at birth, has the same *expected* number of offspring. Of course, some individuals will die sooner and others later and as a result some will wind up with more offspring than others. It is this unevenness and not any differences in Darwinian fitness, that the concentration curve reveals. To the extent a lifetable indicates concentration of reproduction, there is an opportunity for genetic drift, since only a fraction of this generation will produce most of the next generation (Wright 1931, 1938).

Poisson and Binomial Distributions of Reproductive Success

Concentration curves and have-statistics can help ecologists gain insights not only from empirical data and lifetables but also from theoretical models. Consider, for instance, a population of females of some r-selected species. Suppose that all females reproduce once, at exactly the same age, and then die. Further, suppose that each female has n female offspring, where n is some large number, perhaps in the millions, and that each offspring has a small chance $1/n$ of surviving to reproductive age. Thus, each female can expect one reproducing female descendent and the expected population size will remain constant. Note that there are no differences in Darwinian fitness in this model: each female produces exactly the same number of offspring and each of the offspring has exactly the same *chance* of surviving to reproductive age. This model is essentially equivalent to a standard model in evolutionary genetics; Ewens (1979) calls it the Wright-Fischer model because it was implicit in Fischer (1930) and explicit in Wright (1931). The model would be identical to the Wright-Fisher formulation if instead of pertaining to individuals reproducing individuals, the model pertained to genes this generation giving rise to genes next generation.

Although all individuals are equally fit, the randomness of survival will produce an outcome where a few of the females have most of the surviving offspring. The frequency distribution of females by number of surviving offspring, given the assumption of the model, can be approximated by a Poisson distribution. Only about 37 percent of the females will wind up with a single surviving daughter. Another 37 percent will lose all their daughters, about 18 percent will have two

Table 2. Calculations to derive a frequency distribution of individuals by number of offspring from lifetable data.

Age in weeks	Age-specific survival (lx)	Proportion dying in this category	Age-specific natality (mx)	Cumulative natality up to x	Cumulative offspring of those dying in this category
<4.5	-	.13	0.	-	-
4.5	.87	.04	20.0	0.0	10.0
5.5	.83	.02	0.23	20.0	31.5
6.5	.81	.01	15.0	43.0	50.5
7.5	.80	.01	12.5	58.0	64.25
8.5	.79	.02	12.5	70.5	76.75
9.5	.77	.03	14.0	83.0	90.0
10.5	.74	.08	12.5	97.0	103.25
11.5	.66	.07	14.5	109.5	116.75
12.5	.59	.07	11.0	124.0	129.5
13.5	.52	.07	9.5	135.0	139.75
14.5	.45	.09	2.5	144.5	145.75
15.5	.36	.07	2.5	147.0	148.25
16.5	.29	.04	2.5	149.5	150.75
17.5	.25	.06	4.0	152.0	154.0
18.5	.19	.19	1.0	156.0	157.0
		1.00			

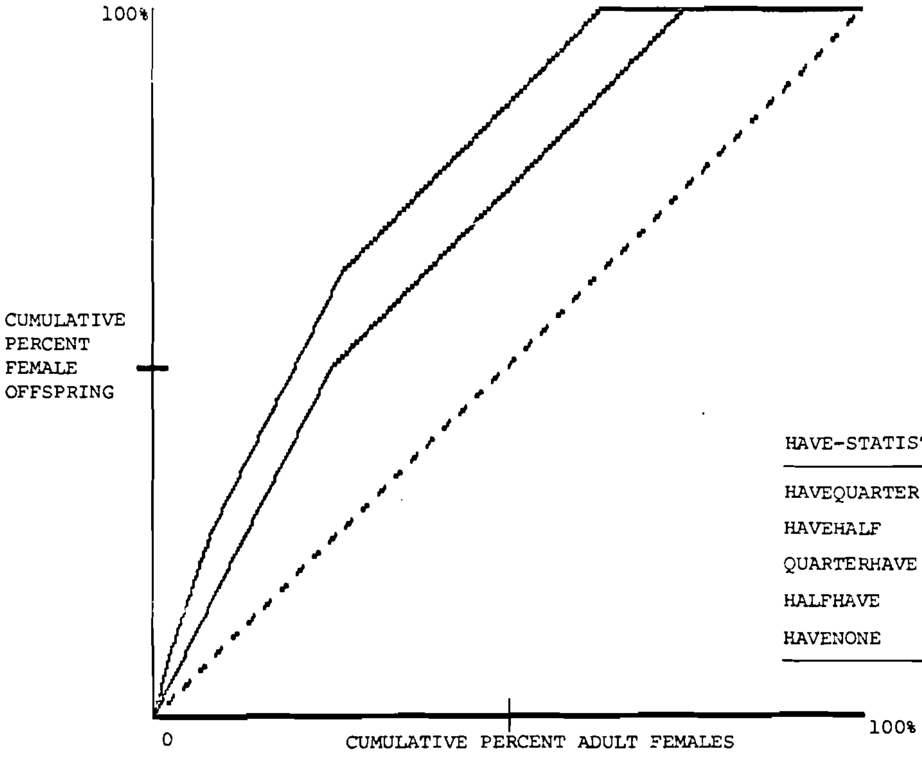
surviving daughters, and the remaining 8 percent of females will have three or more surviving daughters. Given this concentration of reproduction, it turns out that 20 percent of this generation of reproducing females will produce half of the next generation. In sum, even in this case of identical fitness, the havenone is fully 37 percent and the havehalf is only 20 percent. This concentration of reproduction, in a population of finite size, quickens the pace of the evolutionary drift by making the effective population size considerably smaller than the actual population size (Wright 1931, 1938).

The Poisson distribution is an extreme case. For comparison, consider an opposite extreme case of a population of females who, at the age of reproduction, all have exactly 2 female offspring, each of which has a 50 percent chance of surviving to the age of reproduction. As before, the expected number of surviving daughters per female is one, but because of the randomness of survival different females will wind up with different numbers of surviving daughters. In particular, in this model the frequency distribution of females by number of surviving daughters follows a binomial distribution such that 25 percent of the females can be expected to have no surviving daughters, 50 percent to have one, and 25 percent to have two. The 25 percent who have two surviving daughters will account for half of the total number of surviving daughters. Thus, in this model, the havehalf is 25 percent (as opposed to 20 percent in the Poisson model) and the havenone is also 25 percent (as opposed to 37 percent in the Poisson model). What is noteworthy is that the two extreme models, the binomial with two daughters per female and the Poisson with millions of daughters per female, both yield a high concentration of reproduction. Figure 14a plots the two concentration curves.

Consider now the case where a population is not stationary but either growing or contracting. Assume the population is that of an r-selected species, as described above. Suppose, in the case of the growing population, that the chances a newborn will reach reproductive age are $2/n$ (rather than $1/n$): in this case, each female will have an average of 2 surviving daughters. Similarly, in the case of the contracting population, suppose that the chances of survival are $1/2n$, so that each female will have an average of .5 surviving daughters. As before, the frequency distribution of females by number of surviving daughters can be approximated by a Poisson distribution. In the case of a growing population, 26 percent of the females will have half the surviving daughters and 14 percent will have none. For the declining population, the havehalf is 14 percent and the havenone 61 percent. Thus, reproduction is more concentrated in the declining population. This result seems so appropriate that it is worth emphasizing that it is produced by chance alone--there are no Darwinian differentials in fitness at work. Figure 14b displays the relevant concentration curves.

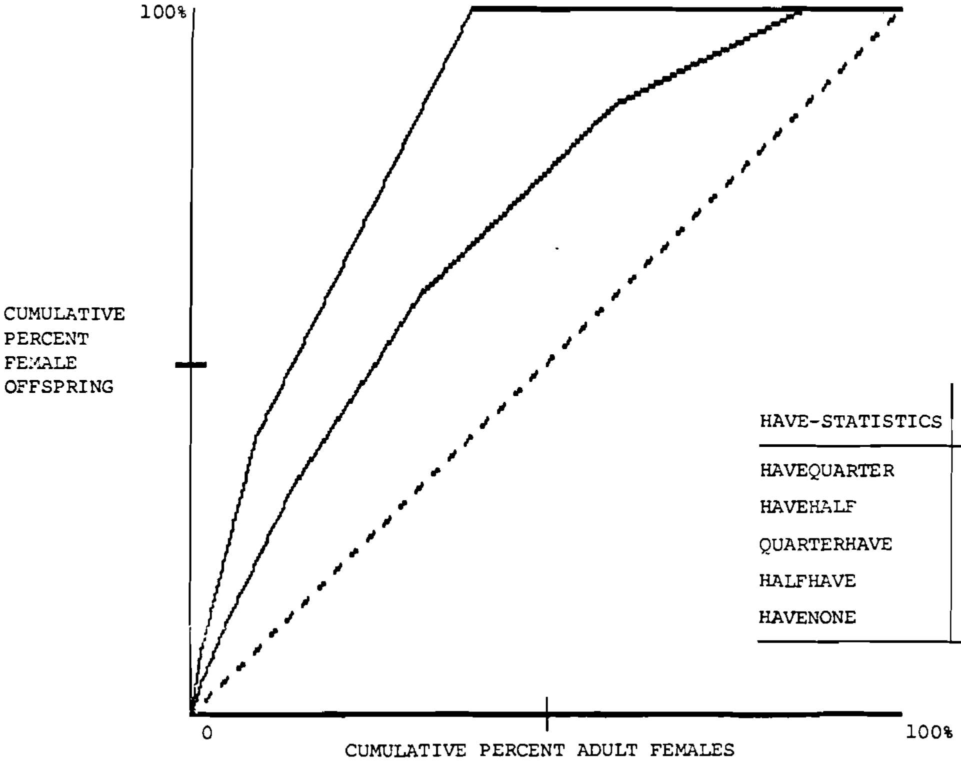
Finally, consider a model where there are differences among individuals in fitness, or expected number of surviving offspring. We make the same assumptions as before concerning an r-selected species except now assume there are two kinds of females, one of which, under current conditions, will have 2 surviving daughters on average and the other of which will only have .5 surviving daughters on average. This differential fitness might be produced by differences in number of offspring or by differences in their survival chances or some combination: all that matters is the resulting differential. Note that this model is essentially a mix of the models of a growing and a declining population described above.

If one-third of the females are in the fit category and two-thirds in the less fit category, then the population will not change in size from this generation to the next: for the population as a whole, there will be an average of one surviving daughter per female. But, assuming heritability of fitness, in the next generation there will be twice as many females in the fit category as in the less fit category: this is Darwinian selection at work. Eventually, if conditions remain the same, the entire population will be the descendants of the fit, although the average number



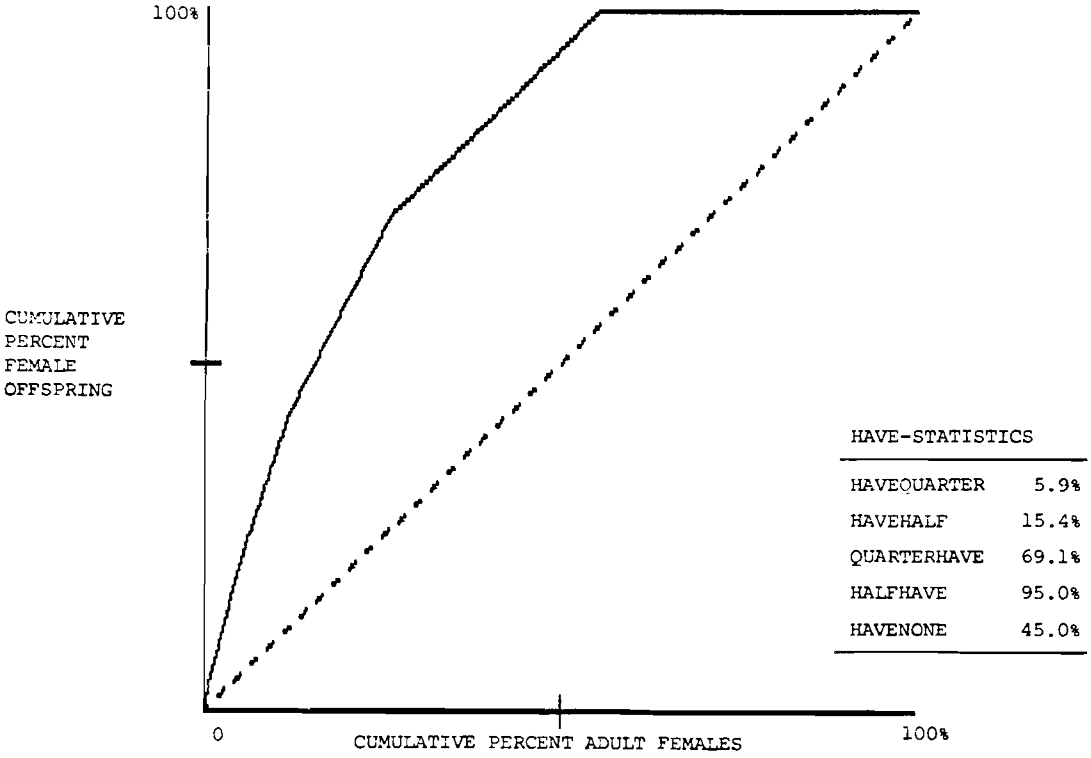
HAVE-STATISTICS	POISSON DISTRIBUTION	BIONOMIAL DISTRIBUTION
HAVEQUARTER	7.6%	12.5%
HAVEHALF	19.8%	25.0%
QUARTERHAVE	60.4%	50.0%
HALFHAVE	86.8%	75.0%
HAVENONE	36.8%	25.0%

Figure 14a: Concentration of lifetime reproductive success in 2 theoretical, stationary, homogeneous populations with survival distributions which are:
- poisson - outer curve
- binomial - inner curve.



HAVE-STATISTICS	DECLINING POPULATION	GROWING POPULATION
HAVEQUARTER	5.4%	10.6%
HAVEHALF	14.4%	26.1%
QUARTERHAVE	71.3%	48.4%
HALFHAVE	100.0%	77.1%
HAVENONE	60.7%	13.5%

Figure 14b: Concentration of lifetime reproductive success in 2 theoretical, homogeneous populations with poisson survival distributions for:
- Growing population - inner curve
- Declining population - outer curve.



HAVE-STATISTICS	
HAVEQUARTER	5.9%
HAVEHALF	15.4%
QUARTERHAVE	69.1%
HALFHAVE	95.0%
HAVENONE	45.0%

Figure 14c: Concentration of lifetime reproductive success in a theoretical, stationary, heterogeneous population with a poisson survival distribution.

of surviving offspring per adult will have to fall from two to one since populations cannot increase in size indefinitely.

As might be expected, reproduction turns out to be somewhat more concentrated in this heterogeneous model than in the earlier homogeneous model, but the difference is small. Some 15 percent of the females have half the surviving daughters (as opposed to 20 percent) and 45 percent have no surviving daughters (as opposed to 37 percent). Drift is still in operation; indeed, at least in this instance, the operation of Darwinian differential fitness somewhat speeds the pace of evolutionary drift. Figure 14c presents the concentration curve for the heterogeneous model.

Conclusion

Since Darwin, a central concern of biologists has been variation, diversity, and inequality, especially concerning reproduction and the differential contribution of individuals to future generations. Some standard tools, most notably frequency distributions and such summary statistics as the coefficient of variation, have been used to study differences among individuals within species. Remarkably, an especially appropriate set of concepts and measures, involving concentration curves and have-statistics, has been largely unexploited. As the examples presented in this paper demonstrate, a frequency distribution tends to visually minimize the importance of those outliers that may be accounting for most of the offspring or other variable of interest; a concentration curve highlights the importance of individuals that are particularly successful or dominant. Similarly, standard measures, like the coefficient of variation, the coefficient of skewness, and entropy, are not only difficult to interpret and intuitively comprehend, but also are often only indirect measures of the quantity of interest, namely the degree of concentration. (A discussion and comparison of various summary statistics is included in Part III of the trilogy of papers of which this paper is the first part.) Concentration curves and the related have-statistics are direct, readily intelligible indicators of concentration.

As illustrated in this paper, concentration curves and have-statistics can be exceptionally helpful in comparative analyses, where the degree of concentration is compared over time, under different conditions, at different stages of life, between sexes, and so on. Greater use of concentration curves in a systematic way for comparative analyses of this sort, concerning, for instance, differences in life patterns between males and females, may give ecologists a perspective and some insights not readily attainable by traditional methods of analysis.

Systematic analysis of diversity, especially in studies of reproductive success, would be greatly facilitated by the availability of appropriate data. In studies of variation among individuals, the most helpful data concern particular, identified individuals, especially when the individuals and their descendents are tracked over their entire lifespans. These data are so difficult to gather and so invaluably useful in analyzing individual differences that it seems particularly unfortunate that most field studies only report summary statistics. A frequency distribution or a concentration curve can be used to summarize data about individuals so economically that there seems to be no excuse for suppressing the information of most direct concern to a life scientist. Lifetable data, although usable, are not as informative as data about particular individuals' lifespans and fertility, because a lifetable averages out any Darwinian differentials in fitness.

A key purpose of a graphical method for displaying quantitative information or of a measure for summarizing such information is to aid comprehension by highlighting the essence of the data and revealing its significance. Such comprehension facilitates insights and stimulates explanatory conjectures and hypotheses. Knowing that the coefficient of variation of the number of hatchlings in some population of bullfrogs is 1.53 may not be as helpful in this regard as knowing that 11 percent of the bullfrogs had half the hatchlings. The mind leaps at this concentration: which 11 percent, what are the characteristics of these highly successful reproducers, what implications does this concentration have for evolution, why and how did such a concentration of reproduction evolve? Concentration curves and have-statistics are useful because they help focus on such relevant and important questions.

Frequency distributions and various summary statistics can also be useful in particular applications; adding concentration curves and have-statistics to them provides life scientists with a richer set of concepts and measures, a more complete set of tools.

REFERENCES

Alker, H.R. (1983) *Mathematics and Politics*. New York: Macmillan.

Arnold, S.J. and M.J. Wade (1984a) On the measurement of natural and sexual selection: Theory. *Evolution* 38:709-719.

Arnold, S.J. and M.J. Wade (1984b) On the measurement of natural and sexual selection: Applications. *Evolution* 38:720-734.

Bateman, A.J. (1948) Intra-sexual selection in Drosophila. *Heredity* 2:349-368.

Clutton-Brock, T.H., F.E. Guinness and S.D. Albon (1982) *Red deer: Behavior and ecology of two sexes*. Chicago: University Chicago Press.

Crow, J.F. (1958) Some possibilities for measuring selection intensities in man. *Human Bioliogy* 30:1-13.

Darwin, C. (1859) *The Origin of Species by Means of Natural Selection*. London: John Murray.

Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Downhower, J.F. and L. Brown (1980) Mate preferences of female mottled sculpins. *Cottus bairdi. Animal Behavior* 28:728-724.

Ewens, W.J. (1979) *Mathematical Population Genetics*. Berlin: Springer Verlag.

Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Oxford: University Press.

Foster, J.E. (1985) Inequality measurement. In *Fair allocation*, edited by H.P. Young. *Proceedings of Symposia in Applied Mathematics*, vol. 33. American Math. Society.

Howard, R.D. (1979) Estimating reproductive success in natural populations. *Amer. Natur*. 114:221-231.

Howard, R.D. (1983) Sexual selection and variation in reproductive success in a long-lived organism. *Amer. Natur*. 122:301-325.

Keyfitz, N. (1968) *Introduction to the Mathematics of Population*. Reading, Mass.: Addison-Wesley.

Kolata, G. (1985) Is the war on cancer being won? *Science* 229:543-544.

Lasswell, H.D. (1958) *Politics: Who Gets What, When, How*? Cleveland: World Publishing Company.

Le Boeuf, B.J. (1974) Male-male competition and reproductive success in elephant seals. *Am. Zool*. 14(1):163-176.

Le Boeuf, B.J. and R.S. Peterson (1969) Social status and mating activity in elephant seals. *Science* 163:91-93.

Lewontin, R.C. (1984) Laws of biology and laws in social science. In *Population and Biology*, edited by N. Keyfitz. Liege, Belgium: Ordina Editions.

Lill, A. (1974) Sexual behavior of the lek-forming white-bearded manakin. *Manacus Manacus Trinitatis Hartert. Z. Tierpsychol.* 36:1-36.

Lorenz, M.O. (1905) Methods of measuring the concentration of wealth. *Journal of American Statistical Association* 9:209-219.

Margalef, D.R. (1958) Information theory in ecology. *General Systems* 3:36-71.

Marshall, A.W. and I. Olkin (1979) *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press.

Maynard Smith, J. (1956) Fertility, mating behaviour and sexual selection in Drosophila subobscura. *Journal of Genetics* 54:261-279.

Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution and Inheritance*. Cambridge, Mass.: Harvard University Press.

Odum, E.P. (1983) *Basic Ecology*. New York: CBS College Publishing.

Orians, G.H. (1969) On the evolution of mating systems in birds and mammals. *Amer. Nat.* 103:589-603.

Patil, G.P. and C. Taillie (1979a) A study of diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. In *Contemporary Quantitative Ecology and Related Econometrics*, edited by G.P. Patil and M.L. Rosenzweig. Fairland, Maryland: International Co-operative Publishing House.

Patil, G.P. and C. Taillie (1979b) An overview of diversity. In *Ecological Diversity in Theory and Practice*, edited by J.F. Grassle *et al.* Fairland, Maryland: International Co-operative Publishing House.

Patil, G.P. and C. Taillie (1982) Diversity as a concept and its measurement. *Journal of American Statistical Association* 77:548-567.

Persons, --. (1908) *Quarterly Journal of Economics*, p. 431.

Pielou, E.C. (1969) *An Introduction to Mathematical Ecology*. New York: Wiley.

Pielou, E.C. (1975) *Ecological Diversity*. New York: Wiley.

Simpson, E.H. (1949) Measurement of diversity. *Nature* 163:688.

Taillie, C. (1979) Species equitability: A comparative approach. In *Ecological Diversity in Theory and Practice*, edited by J.F. Grassle *et al.* Fairland, Maryland: International Co-operative Publishing House.

Theil, H. (1972) *Statistical Decomposition Analysis*. Amsterdam: North Holland.

Wade, M.J. (1979) Sexual selection and variance in reproductive success. *Am. Nat.* 114:742-746.

Whittaker, R.H. (1965) Dominance and diversity in land plant communities. *Science* 147:250-260.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16:97-159.

Wright, S. (1938) Size of population and breeding structure in relation to evolution. *Science* 87:430-1.