Statistics Helps Breeding Farm Animals of High Productivity

BU-1372-M October, 1996

Shayle R. Searle
Biometrics Unit
Cornell University
Ithaca, N.Y., 14853

ABSTRACT

This is a brief, non-mathematical review article of some uses of statistics in animal breeding, invited by the editor of *Chance*.

1 A FARMER'S OBJECTIVE

Farmers the world over want to own farm animals that yield large amounts of their marketable products, be they dairy cows producing milk, beef cattle producing meat, pigs having large litters and/or developing meat carcasses, sheep producing lambs for meat and fleeces for wool, or poultry laying lots of eggs or rapidly growing into tender, plump broilers.

In looking to the future, one farming objective is to have animlas that yield at least as much as (preferably more than) those presently owned. Two major contributions in this objective are breeding and farm management. The latter consists of numerous environmental factors including, for example, animal health and nutrition, and facilities for animal maintenance such as pastures, feed additives, barns and so on. In contrast is breeding, the general approach to which is to mate animals of high yield in the hope of (and, over the long term) getting offspring that are themselves high yielders. However, not every such mating gives high-yielding offspring. For example, not all the matings to Triple Crown winner Secretariat have produced consistent winners on the race track.

In all animal productivity there is variability among yields. Dairy cows do not all yield the same amount of milk each year; even the same cow has different milk yields from year to year. And, of course, hens vary in their egg productivity.

This variability, which is due partly to genetics and partly to environment, is referred to as variance by statisticians. It gives farmers the opportunity to select from amongst their animals those of higher productivity from which to breed the next generation. However, high productivity has to be thought about very carefully because an animal which has a high yield this year may not do so next year. And vice versa.

2 GENETICS AND ENVIRONMENT

Generally speaking, yield depends on both genetics and environment, and both need to be taken into account when making decisions about what animals to mate for breeding the next generation; or, equivalently, when every animal in one's herd of cows or flock of ewes gets mated every year, in making decisions about which of the newly-borns will be selected to remain on the farm as future producers.

At the time of birth, the genetic make-up of a newly-born's potential productivity is fully determined and in wanting that productivity to be high we want the genetic contribution thereto to be high also. Since that genetic contribution comes from the genetic make-up of each of the parents, the question then arises as to how can we estimate from production records the genetic value (which cannot be measured directly) of animals used as parents, so that they can be ranked, to allow us to retain offspring from only the highest ranking ones. This is where the use of statistics enters the picture. We consider this problem in terms of milk yields of dairy cows.

3 USING AN ANIMAL'S OWN RECORDS

A cow is milked for approximately 305 days after calving, a timespan known as a lactation. Following that, after some 2-3 months of not being milked, it calves again and the next lactation begins. Throughout a lactation, the periodic (usually every 1, 2 or 3 months) weighing of each cow's daily milk yield provides its (estimated) total lactation milk weight. All such weights are then adjusted for age of cow, the time of year when lactation began, and the number of lactations prior to the one at hand. This adjusted weight, which we call yield, is then used to compare cows of the same breed (since, for example, Holsteins give more milk than Jerseys) within a herd.

Estimating a cow's genetic value begins by averaging several (n, say) lactation yields, denoting that average by $\bar{y}$. Now of all the cows in a herd, there will probably be several that have essentially the same $\bar{y}$. And even if that is not so, we can imagine a hypothetical population of such cows. But those cows, despite all having the same $\bar{y}$, will not all have the same genetic value. The best we can do therefore, for a cow with given $\bar{y}$, is to estimate what we think is the average genetic value of the population of cows that do or might have the same observed $\bar{y}$. Actually, from $\bar{y}$ we subtract the average yield of the whole herd, call it μ, and then estimate the cow's merit c by

$$\hat{c} = \mu + k(\bar{y} - \mu).$$

The question then is, what is k? It is a weighting factor somewhat of the same nature as the well-known "regression towards the mean." It is, indeed

$$k = \frac{nh^2}{1 + (n-1)r}$$

where $\bar{y}$ is the average of n lactation yields, h^2 (historically denoted h^2 and not h) is the fraction of the total variance among lactation yields that is due to genetics variance, and r is the intra-cow correlation of lactation yields, i.e., the correlation between records of the same cow. In animal breeding parlance, h^2 is called heritability and r is called repeatability. These two parameters by their definition, are where the variability among lactation records due to genetics and to environment enters into the estimation of a cow's genetic value. Those parameters can be estimated as ratios of estimated variance components (or sums thereof), and this is where statistics plays a big part. Methods of estimating variance components is a statisticsal procedure that has commanded considerable space in the statistical research literature of the last forty-five years. And it is a procedure that received a big boost from animal breeding considerations of the nature described here.

For dairy cow records, $h^2 = .25$ and $r = .55$ are oft-found values; they reduce k to

$$k = \frac{5n}{9n + 11}$$

Thus for $n = 1$, 2, 3 and 4, the values of k are .25, .34, .39 and .42, respectively with the maximum possible value (if n were infinite) being .55. Using these values in $\hat{c}$ we can then rank cows according to their $\hat{c}$-values and those with the highest such values can be considered the best for breeding the next generation. What is nice about this is that theory tells us that in ranking the cows according to their $\hat{c}$-values we are maximizing the probability of correctly ranking the cows according to their true genetic values. Searle *et al.* (1992, p. 268) have references to the proof of this result.

A measure of the reliability of $\hat{c}$ as an estimate of c is the correlation between $\hat{c}$ and c. This turns out to be $\sqrt{k}$ which for $h^2 = .25$, $r = .55$ and $n = 1$, 2, 3 and 4, is .50, .58, .62 and .65 respectively, so indicating that 3 or 4 lactation yields give a somewhat more reliable estimate of genetic value than does a single yield.

Although the preceding description is in terms of cows producing milk, it applies equally as well, for example to sheep (ewes or rams) producing wool, or to sows producing successive litters of piglets. But it does not apply, for example, to beef cattle producing prime beef (because they have to be slaughtered to produce beef), nor to dairy bulls (because they do not give milk). In these cases we have to estimate the genetic merit of a bull, say, from yields of its ancestors, or of its offspring. Relying on ancestor records is not very satisfactory bcause for dairy bulls, for example, only the female ancestors can have milk yields.

On the other hand, with the now widespread use of artificial insemination, it is easy for a bull to have large numbers of daughters (100,000 and more) with milk records, and these are reliable for assessing the value of a bull. This situation is extremely valuable to the dairy farming industry round the world. A description follows.

4 USING OFFSPRING RECORDS

A vital feature of today's dairy industry is the existence of artificial breeding companies. Their purpose is to own a smallish number of bulls (say 10-20) of high genetic merit with the objective of using them for breeding daughters that are high yielding milk producers. Semen is collected from these high merit bulls, it is extensively diluted and frozen in ampules (or 'straws' as they are called in the industry), and then distributed and inseminated into cows in a wide variety of farms. By this means, because the bulls so used have been highly selected, the per cow milk-producing capacity of farms using this service has increased enormously over the last 30 years. For example, In N.Y. State with its approximately 750,000 dairy cows, it is figured that artificial breeding has led to an increase of 100 lb. of milk per cow per year for the last 40 years. (There has been a commensurate increase of 200 lbs. due to management.)

The secret to this success story is the selecting of the high merit bulls for heavy service (e.g., 75,000 matings per year). An artificial breeding company achieves this selection by each year buying a large number of young bulls (say 100-200) and selecting from them those few that have the highest estimated genetic values for siring high-producing daughters. This is done by seeing that semen from each young bull goes to some 300 cows spread over many different farms, from which, four years later, there will be around 50 daughters of the bull that have a complete lactation record available. (300 cows give birth to some 150 cow calves, not all of which are retained in their herds, and of those that are some will die or themselves not get in calf. Four years are needed: approximately one for the gestation of a cow inseminated with a young bull's semen; one for its daughter calf to become a year old and itself be mated, one for that daughter's gestation, and finally one for the now 2-year-old daughter to complete a lactation yield.)

It is these 50 daughter lactation yields, for each of the many young bulls, that are used to estimate the genetic value of each young bull. The bulls are then ranked according to those estimates and the highest 2, 3 or 4, or however many are needed to sustain the company's cadre of high merit bulls, are retained. The remainder are disposed of (possibly fattened for bull beef, which commands a premium in some markets).

Again, it is statistical methods that are used to estimate a bull's genetic merit from its 50 or so 2-year-old daughters' lactation records. Let the average of those m records, say, be $\overline{x}$. Then the estimate of the sire vaue s is just like preceding estimation of a cow's value,

$$\hat{s} = \mu^* + t(\overline{x} - \mu^*),$$

except the details are different: μ^* is some appropriate population average and

$$t = 2mh^2/[4 + (m-1)h^2]$$

where h^2, called heritability, is a ratio of the genetic variability to the total variability in milk records. Again, it is the statistical methods of estimating causes of variability, methods known as variance components estimation, that are useful for estimating the variances needed for h^2.

A not untypical value for h is 0.25 which reduces t to

$$t = \frac{2m}{m+15}$$

which, for m =50, 75, 100 and 200 has values t = 1.53, 1.66, 1.74 and 1.86 respectively. At first thought these values being greater than 1.00 might seem surprising. But it is not, because an offspring is genetically only half of its sire, and so estimating sire values from offspring records would seem to demand, at least for an infinite number of offspring, a multiplyer of 2; and this is so: t tends to 2 as m gets larger and larger.

The reliability of $\hat{s}$ (correlation of $\hat{s}$ with s) is $\sqrt{t/2}$ which, for $m = 50$, 75, 100 and 200 is .87, .91, .93 and .96 respectively, its limit being 1.00 for m being infinite.

One well might ask at least two questions about the sire selection process just described. First is "How are the young bulls selected for purchase?" One answer is that ancestor records such as those of dam and grandam can be used to assess genetic merit; also records of half-sisters having the same sire, or of full sisters if available. A second question is "how are the cows selected to which the young bulls will be mated?" Very broadly, is the answer, so that they can be considered as a random sample of all cows to which bulls might be mated.

Whatever the details it is clear from numerous sources that this kind of sire selection program for artificial breeding has been very successful. In New York State an increase of 4,000 lb. of milk per cow, over the last 40 years is attributable to such a program. And it is even more in other parts of the world: New Zealand for example.

5 COMMENTS

The preceding descriptions have been designed to give a general impression of how animal breeding for improved production utilizes statistics. Not only have all details of the necessary statistical methods been omitted, but so also have numerous practical details of implementing selection procedures on individual farms or in artificial breeding companies. Readers seeking further information could refer to numerous books of which the following four might be found suitable

Schmidt G.H., Van Vleck L.D. and Hutjens M.F. (1988) *Principles of Dairy Science,* Prentice-Hall, Englewood Cliffs, N.J.

Searle S.R. (1987) *Linear Models for Unbalanced Data,* Wiley and Sons, New York, N.Y.

Searle, S.R. Casella G. and McCulloch C.E. (1992) *Variance Components,* Wiley and Sons, New York, N.Y.

Van Vleck L.D., Pollak E.J. and Oltenacu E.A.B. (1987) *Genetics for the Animal Sciences,* W.H. Freeman and Co., New York, N.Y.