

Estimating the Size of the Population

Often, we estimate means, totals, and proportions, assuming that the population size is either known or so large that it could be ignored. Frequently, however, the population size is not known and is important to the goals of the study. In fact, in some studies, estimation of the population size is the main goal. The maintenance of wildlife populations depends crucially on accurate estimates of population sizes.

Direct Sampling

One common method for estimating the size of a wildlife population is *direct sampling*. This procedure entails drawing a random sample from a wildlife population of interest, tagging each animal sampled, and returning the tagged animals to the population. At a later date, another random sample of a fixed size n is drawn from the same population, and the number s of tagged animals is observed. If N represents the total population size, t represents the number of animals tagged in the initial sample, and p represents the proportion of tagged animals in the population, then $\frac{t}{N} = p$. Also, we expect to find approximately the same proportion (p) of the sample of size n tagged as well. So $p \approx \frac{s}{n}$. This gives us a way to estimate the size of the population N , since $\frac{s}{n} \approx \frac{t}{N}$. Solving for N provides an estimator for N ,

$$\hat{N} = \frac{nt}{s}.$$

The approximate estimated variance of \hat{N} is

$$\hat{V}(\hat{N}) = \frac{t^2 n(n-s)}{s^3}.$$

Notice that we have serious problems when s is zero, and a large variance when s is small.

As an example, suppose we initially capture and tag 200 fish in a lake. Later, we capture 100 fish, of which 32 were tagged. So $t = 200$, $n = 100$, and $s = 32$. Then our estimate of N is

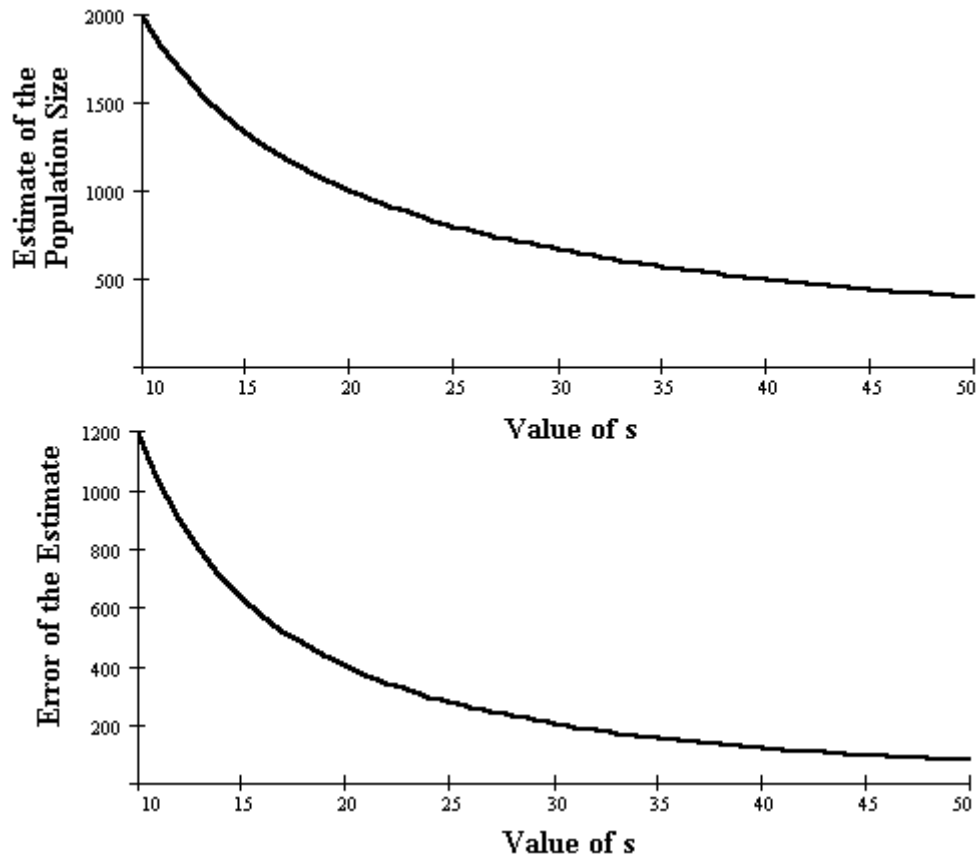
$$\hat{N} = \frac{nt}{s} = \frac{100(200)}{32} = 625 \text{ fish.}$$

Also, we approximate the variance with

$$\hat{V}(\hat{N}) = \frac{t^2 n(n-s)}{s^3} = \frac{(200)^2 (100)(100-32)}{32^3} = 8301.$$

The margin of error is $2\sqrt{\hat{V}(\hat{N})} = 2\sqrt{8301} = 182$. Our estimate of the number of fish in the Lake is between 443 and 807 fish.

The graphs below illustrate how sensitive are both the estimate of N and the margin of error when s is small. If s is less than 4, the error of the estimate is larger than the estimate for these values of n and t .



Inverse Sampling

We can get around the problem of having a small value of s by sampling until we have a pre-specified value of s . For example, we could fish until we have caught 50 of the tagged fish. This technique is called *inverse sampling*. That is, we sample until a fixed number of tagged animals, s , is observed. Using this procedure, we can also obtain an estimate of N , the total population size by computing $N = \frac{nt}{s}$. This is the same computation as before, only

s is fixed and n is random. This changes the variance of \hat{N} . The estimated variance of \hat{N} is

$$V(\hat{N}) = \frac{t^2 n(n-s)}{s^2 (s+1)}.$$

This variance estimate is almost the same as before, but it can be considered as a function of n , rather than s . We no longer have to worry about a small s , but this procedure may take much longer and be more expensive, since we do not know how long we need to continue the recapturing process before the pre-set value of s is achieved.

Example Consider our earlier example in which we initially captured and tagged 200 fish in a lake. Later, we fished until we had captured 50 fish that had

previously been tagged. This required us to catch 162 fish. So $t = 200$, $n = 162$, and $s = 50$. Then our estimate of N is

$$\hat{N} = \frac{nt}{s} = \frac{162(200)}{50} = 648 \text{ fish.}$$

We approximate the variance with

$$\hat{V}(\hat{N}) = \frac{t^2 n(n-s)}{s^2(s+1)} = \frac{(200)^2 (162)(162-50)}{50^2(51)} = 5692.$$

The margin of error is $2\sqrt{\hat{V}(\hat{N})} = 2\sqrt{5692} = 151$. Our margin of error is smaller since we forced a larger value of s , but it required more resources to catch the extra 62 fish.

Another method for computing the estimated interval is to find a confidence interval on the proportion $\hat{p} = \frac{s}{n}$ and use it to create the interval for N algebraically. In our first example, we have $\frac{s}{n} = \frac{32}{100} = 0.32$. A 95% confidence interval for this proportion is

$$0.32 \pm \frac{1.96\sqrt{0.32(0.68)}}{\sqrt{100}} = 0.32 \pm 0.09 \text{ or } (0.23, 0.41).$$

Now, we have $\hat{N} = \frac{n}{s}t$, so our estimate is $\hat{N} \approx \frac{t}{(\frac{s}{n})} = \frac{200}{0.32} = 625$ fish. An interval estimate can

be derived using the two extremes of the interval (0.23, 0.41). So $\hat{N} \approx \frac{t}{(\frac{s}{n})} = \frac{200}{0.23} = 870$ and

$\hat{N} \approx \frac{t}{(\frac{s}{n})} = \frac{200}{0.41} = 488$. Our estimate then is between 488 and 870 fish. Notice that the point estimate 625 is not in the center of the interval (488, 870).

Experimental Design for Capture-Recapture

There are two factors, t and n , that influence the variability of the estimate of N when using capture/recapture. A common question about capture recapture is, "Is it better to mark more fish initially or is it better to take a larger sample in the recapture phase?" The question is really about where to put your energy and resources. The recapture phase can be repeated many times and the resulting estimates of N compared, perhaps in a stem-leaf plot. One could then vary the number of tagged animals and repeat the process to see how the variability of the estimates depends on t . One could also vary the size of the second capture to see how the variability of the estimates depends on n .

Below are box-plots comparing 100 estimates of $N = 1000$ using either $t = 100$ or $t = 200$ and either $n = 60$ or $n = 120$. From the boxplots you can see that changing t from 100

to 200 has a greater effect on reducing the variability than does changing n from 60 to 120. Greater benefits are achieved when more effort is put on the initial sample to be tagged. Note that when both t and n are small, small values of s were more often generated producing large estimates of N .

