Maddox Johnston
Math 531: Dr. James Young

Case Study: Vincent, Jean-Paul, et al. "Testing Density Estimators on a Fallow Deer Population of Known Size."

Summary

This section aims to summarize the methods and results of this study done by four researchers from the French Institut de Recherche sur les Grands Mammiferés.

Background

This research is unique in the fact that it tests common population and density estimators on an enclosed ungulate population of known size: a population of Fallow Deer (*Dama dama*) freely roaming a 321-acre woodland enclosed by deer proof fence in Northeast France. Data reported is from April 1992 to May 1994, broken up seasonally. The researchers were able to obtain seasonal population/density estimates using *Kilometric Index, Petersen-Lincoln Estimate*, and *Strip Transect* methods, and then evaluate causes of bias due to the population being of known size.

Methods

Before being able to test these estimators, researchers were tasked with marking a large subset of the deer population. Marking occurred as "often as possible, during winter, spring and autumn (pre-rut), with the objective of marking as many deer as possible, regardless of sex or age" (Vincent et al., 1996, p.19). They also took this opportunity to note the age and sex of all marked individuals, in order to accurately construct the population distribution. Birth rate and death rate were monitored (all carcasses would be apparent), but "natural mortality was almost non-existent" (Vincent et al., 1996, p.20) over the course of the study. Eventually, all individuals were marked.

Kilometric Index is a commonly used density estimator for large animals, defined as the number of individuals spotted per kilometer travelled by an observer. These observations are done in a standardized way; they are carried out during the same time of day and same period each season. A 4.1 mile transect length was deemed suitable due to the homogeneity of the habitat. Researchers later fit a linear regression on the Kilometric Index (see below: here population size (N) is the independent variable), and compared percent change of the Kilometric Index with the known population percent change.

$$\widehat{N} = \beta_0 + \beta_1 N$$

***Equation 1**:* A simple linear regression for estimating population size.

The *Petersen-Lincoln Estimate* is a common population estimator involving populations with a subset of marked individuals.

$$\widehat{N} = M(n+1)/(m+1)$$

***Equation 2**:* The Petersen-Lincoln population estimator.

Here M is the known number of total marked deer in the population, m is the number of marked deer observed, and n is the number of total deer observed. Observations were carried out in a similar manner as for the Kilometric Index; researchers walked transects daily during each 2-3 week observation period per season. It is worth noting that the Petersen-Lincoln Estimate could not be calculated in the last two seasons of the study, since all individuals were marked.

Bias was expressed using percent relative bias, noticing the significance of the bias follows a *Z*-distribution, defined as below:

$$Z = (\widehat{N} - N)/(\widehat{SE}(\widehat{N})$$

Equation 3*:* The standardized difference of the population estimates and true values.

Where the Standard Error is defined as:

$$\widehat{SE}(N) = \sqrt{[M^2(n+1)(n - m)/(m+1)^2(m+2)}$$

Equation 4*:* Standard Error of the Petersen-Lincoln population estimator.

Here, if there was no bias present, Z would become approximately standard normal, while significant bias would result in large Z values.

The *Strip Transect* method consists of creating a set of points in the habitat where the maximum perpendicular distance an animal is capable of being visible from, w, is known[1], thus creating “strips” of width $2w$. The researchers created 71 “transect blocks” in this fashion, with the sum of all these blocks equaling the total study area A. Population size was then estimated from the formula below, where n is the number of individuals observed, L the length of the transect block, and A and w are as described before.

$$\widehat{N} = An/\sum(2Lw)$$

Equation 5*:* The Strip-Transect population estimator.

Researchers similarly used percent relative bias and a *Z*-distribution to evaluate the significance of the bias for this estimator.

Results

Tables displaying the results of the various estimators are available in the Appendix. The Linear Regression derived from the results of the Kilometric Index was found to be significant[2], and described the equation $y = 1.058 + 0.071x$. Similarly, the percent change of the Kilometric Index mirrored the percent change of the population in all seasons besides Summer 1993, due to “difficulty of observation, particularly of young fawns, in dense summer cover, leading to an underestimation of population size” (Vincent et al., 1996, p.25).

---

[1] We note that w varies seasonally due to vegetation limiting visibility; each season’s Strip-Transect calculations use a different value of w.

[2] We have a R^2 value of 0.75, and a p-value of 0.012

Similarly, the Strip Transect method suffered from poor visibility due to Summer vegetation, so the method was only used during the Spring and Winter seasons. However, in the months where the method was possible, the Z-scores indicate insignificant bias in all but one case (Spring 1992). This population estimate's corresponding regression equation was also found to be significant[3], and modeled by $y = -4.819 + 0.981x$. The predicted percent change in population obtained from this population estimate was also in concordance with the actual percent change in population size[4].

Although the Petersen-Lincoln Estimate always yielded percent relative bias below 10%, the population estimates were rarely accurate. The estimates were not consistently skewed in one direction, although a weak seasonal trend related to overestimation in Winter/Spring and underestimation in Summer/Fall did arise. Similarly, although the predicted percent change in population size obtained from the Petersen-Lincoln Estimate often agreed with the actual percent change in cardinality, it rarely agreed in size. For example, between Spring and Autumn 1992, a 21.8% percent change was estimated, but a 7.8% change occurred, and between Autumn and Winter 1992 a 17.8% percent change was estimated despite a constant population.

Since all marked individuals were of known age and sex, this study also leads to interesting insight regarding the varying observability of different sections of the population structure. Chi-squared tests show departure from the expected age/sex distribution in the Summer of 1993 and the Winter of 1994, suggesting that the assumption of equal observability does not hold[5].

Further evidence against the assumption of equal observability is found in the bias present in in August 1992. Throughout 285 individual observations divided among 50 distinct animals, we would expect an average of 5.7 sightings per individual; however, the results were far different in practice, with 6 animals only being observed twice, and one especially active deer observed 13 separate times. This suggests that certain individuals were inherently more likely to be observed during this period.

Previous researchers noted that bias in the Petersen-Lincoln estimate could be caused by not marking a large enough subset of the population, however the most unbiased estimate in this study was that when which the lowest percentage of the population was marked. Indeed, most bias in this study can be explained by ecological factors related to observability. The inaccurate percent change estimate yielded from the Petersen-Lincoln method despite a stable population between Fall 1992 and Winter 1993 seems closely tied in with the mating season (referred to as "the rut") disrupting behaviors and social structure. Similarly, the overrepresentation of males and underrepresentation of females in Winter 1994 can be explained by the tendency for males to form "large bachelor groups, while female group size is smaller" (Vincent et al., 1996, p.26) in colder months.

[3] We have a R^2 value of 0.96, and a p-value of 0.003
[4] We have a R^2 value of 0.96, and a p-value of 0.014
[5] Residual analysis shows that in Summer 1993, fawns of both sexes were underrepresented in the observed distribution. Similarly, in Winter 1994, there was an underrepresentation of Females and an overrepresentation of Males across many different age classes.

Critique

By creatively carrying out this study on a population of known size and structure, the researchers were able to evaluate and demystify some theorized causes of bias for wildlife population/density estimation. However, the researchers were ultimately unable to escape some of this unavoidable bias that seems inherent to any capture/recapture study in practice.

This study took place in a virtually homogenous wooded enclosure, dominated by various species of large trees and thick brush-similar to natural habitat. While this aspect was helpful toward equal chances of observability in differing areas, it caused spotting the animals to be quite difficult itself in summer months when the brush becomes quite overgrown and thick, providing natural cover for more wary individuals. Perhaps some measures could have been taken from preventing this occurrence, without disturbing the natural habitat too much.

It was noted above that there during certain periods, the actual distribution of individuals observed among 50 distinct animals was far from what we would expect. This phenomenon can be partially explained by inevitable behavioral changes during the rut; however, the marking process itself could cause long term stress in some animals (Jessup, 2018). Marking was done "by darting with the anesthetic Zoletil from a 4-wheel drive vehicle" (Vincent et al., 1996, p.19), and researchers remark that there was some difficulty successfully darting certain individuals, in addition to two fatal accidents occurring during the marking process.

The constant marking process was necessary to obtain age/sex data for all individuals in the population (in turn allowing population structure analysis). It is worth noting that beginning population was disproportionately female, because of a population cull before the study began. Eventually the entire population was marked, and although this eliminated the chance for marked individuals to behave differently than unmarked individuals, it made it impossible to use the Petersen-Lincoln method. Similarly, most studies involving Kilometric Index and Strip Transect methods of population/density estimation do not involve populations of marked individuals, that is, individuals who have been perturbed by researchers in the past.

Throughout the course of the study, the population size was obviously fluctuating; a nonconstant population is desired since much of the practical value in population estimation comes from tracking population changes (in terms of direction and amplitude) rather than obtaining an accurate number of individuals. However, coupled with the fact that data for certain statistics was unavailable in certain years, this experiment compares different statistics over non standardized time periods, and uses the different statistics to estimate populations of differing sizes. This causes issues in testing the estimators against each other. For example, we have no way to know how the Petersen-Lincoln estimate would perform using a small population size, or how the Strip-Transect method would compare with the Petersen-Lincoln estimates for Summer months. This lack of comparability between the estimators for certain time periods slightly detracts from the usefulness of this study.

I was surprised by the fact that despite all the causes of bias in the experiment, and discussion of bias in the report and related literature, bias did not seem to be the issue at all for the Petersen-Lincoln method; rather, it was the <u>accuracy</u> itself of the estimator. For example, in Autumn 1992 there were $N = 95$ individuals, and the Petersen-Lincoln estimator yielded an estimated population of $\widehat{N} = 87.3$ with a 95% confidence interval of $[81.6, 93.0]$; similarly, just one season later in Winter 1993 there were still $N = 95$ individuals, but this time the Petersen-

Lincoln estimator yielded an estimated population of $\widehat{N}$ = 102.8 with a 95% C.I. of [95.6, 110.0]. We observe that on both occasions the actual value was not contained in the 95% C.I., and in the Fall we had an overestimation while in the Winter we had an underestimation. While we are not overconcerned with having an accurate estimator if it is precise, we have neither in this instance. Similarly, a linear regression of the rate of change Petersen-Lincoln estimator on the actual population rate of change was insignificant[6].

The $(n+1)/(m+1)$[7] part of the equation for the "closed population" form of the Petersen-Lincoln estimate is meant to scale the total number of marked individuals, M, toward N. The fact that we had overestimates and underestimates means that this scaling factor was not consistently too large or too small. Further, the fluctuations between overestimation and underestimation were random, despite the percentage of marked individuals-the reciprocal of what this scaling factor attempts to capture- consistently increasing throughout the study. I would love to see data for more than 5 seasons to further elucidate this trend, and perhaps reveal what the desirable percentage of marked individuals should be, sans preoccupation with the bias that has been suggested at lower percentages.

One aspect of this study I found particularly interesting was the ability to quantify the varying observability of different segments of the population structure. The researchers remark "Several authors have concluded that bias may be due in part to violation of the assumption of equal observability among animals" (Vincent et al., 1996, p.24). Indeed, this was the case during many periods of the study, as touched on earlier; male deer becomes very erratic and easily perturbed during the rut, and later become gregarious (and thus easily observed) in the Winter post-rut while females tend to go solo. Similarly, all seasonal behavioral factors aside, certain individuals seem more prone to being observed, although it is interesting to note that a singular Male with a conspicuous all-White coat was not consistently under or over-observed (Vincent et al., 1996, p.24).

Spring 1992	0.479
Autumn 1992	0.475
Winter 1993	0.482
Spring 1994	0.360

***Table 1**:* Individual encounter rate during a Strip Transect

I also found it interesting how the probability for an individual to be observed during any Strip Transect was very similar throughout the course of the study, as shown in the table on the left, however these values were consistently well below ratio of area covered by the transect and the total area of the enclosure, 0.644. This led to the resulting Strip Transect population estimator to be a consistent underestimation.

Perhaps this has something to do with the fact that the process of analyzing visibility distance was rather rudimentary, consisting of a researcher holding a brown-painted board continuously walking backward from an observer until the observer can no longer see the board. I would hypothesize that it is easier to detect an object when you already know the object is present, and the paper does not mention accounting for this fact.

[6] We have a R^2 value of 0.25, and a p-value of 0.294.
[7] Here *n* is the number of individuals observed, and *m* is the number of marked individuals observed.

Setting aside the periods clearly tarnished by unequal observability, there were certain periods when the assumption of equal observability among differing age/sex classes was not found to be violated, and there was still significant bias in the experiment. This bias would certainly be magnified under non-standardized “wild” conditions where these wildlife population/density estimators would be used in practice.

Further Research

Although this study is somewhat dated (began in 1992), many methods carried out by the researchers remain relevant today. However, this is not to say that many of these methods could not be improved with current technology.

One major issue encountered in this study was regarding observability and detectability of the animals. In 2018, Alonso et al did a Population Estimation study on a Bobcat population where they assessed the viability of humans and cameras in resighting the Bobcats, finding that a hybrid model using a combination of both humans and cameras was most effective (Alonso et al., 2018). The use of cameras in place or in addition to humans for the Strip-Transect method could prove effective in reducing some of the bias found regarding the detectability of the animals.

Similarly, GPS tracking could also be useful here. If the researchers were able to attach a small GPS tracking device to the animals during the tagging process, by the time all the individuals in the population were marked, we would be able to better evaluate many causes of bias in the experiment. For example, the researchers would be able to know exactly how many individuals were in the Strip-Transect area at any given time and deduce if the discrepancy causing underestimation is due to detectability.

Some discussion regarding the usefulness of the Peterson-Lincoln estimator under conditions where the probability of observing (recapturing) an animal varies by the individual is warranted. Otis et al. recommends using an interpolating jackknife estimator in this instance to reduce bias; here we would treat each age/sex class as its own species. Further, Otis created the CAPTURE model selection program for this closed population estimation problem, to account for the bias caused from seasonality, behavioral changes due to the marking process, and heterogeneity. Several of these models included in the CAPTURE program have outperformed the Petersen-Lincoln estimate in similar studies (Menkens & Anderson, 1988); this is to say that we could perhaps find a better model for the Mark-Recapture data, if it was available.

As I remarked in the Critique section, I would be interested to see how the different estimators compared against each other on a standardized time scale, and with more data. This would allow some insight into to seasonality and the fluctuation between underestimations and overestimations we observed with the Petersen-Lincoln statistic.

Appendix

Table 1. Comparison of known population size with the kilometric index (deer counted/km) during 7 periods in an enclosed population of fallow deer in France, spring 1992–spring 1994.

	Known population		Kilometric index			
Period	N	% change	Low CL	KI (with 95% limits)	High CL	% change
Spring 1992	78		4.98	6.08	7.17	
		21.8				50.3
Autumn 1992	95		7.23	9.14	11.10	
		0				−2.1
Winter 1993	95		6.96	8.95	10.90	
		−10.5				−10.6
Spring 1993	85		6.11	8.00	9.90	
		41.2				−7.2
Summer 1993	120		6.62	7.42	8.29	
Removal	↔	−68.3		↔		−56.6
Winter 1994	38		2.32	3.22	4.12	
		0				2.8
Spring 1994	38		2.58	3.31	1.03	

Table 2. Comparison of known population size with the strip-transect estimate during 5 periods in an enclosed population of fallow deer in France, spring 1992–spring 1994.

	Known population		Population estimate						
Period	N	% change	Low CL	$\hat{N}$ (with 95% limits)	High CL	% change	PRB	Z	P
Spring 1992	78		50.8	62.8	74.7		−19.5	−2.80	0.005
		21.8				46.3			
Winter 1993	95		71.4	91.8	111.8		−3.3	−0.35	0.726
		−10.5				−10.6			
Spring 1993	85		62.7	82.1	101.5		−3.4	−0.33	0.741
		−55.3				−59.8			
Winter 1994	38		23.8	33.0	42.3		−13.1	−1.17	0.242
		0				2.8			
Spring 1994	38		26.5	34.0	41.4		−10.7	−1.17	0.242

PRB: percent relative bias = $100 \cdot (\hat{N} - N)/N$.
Z: standardized difference = $(\hat{N} - N)/SE(\hat{N})$.

Table 3. Comparison of known population size with the Petersen-Lincoln estimates during 5 periods in an enclosed population of fallow deer in France, spring 1992–summer 1993.

	Known population		Population estimate						
Period	N	% change	Low CL	$\hat{N}$ (with 95% limits)	High CL	% change	PRB	Z	P
Spring 1992	78		73.7	81.0	88.3		3.85	0.81	0.418
		21.8				7.8			
Autumn 1992	95		81.6	87.3	93.0		−8.11	−2.66	0.007
		0				17.8			
Winter 1993	95		95.6	102.8	110.0		8.21	2.13	0.033
		−10.5				−10.6			
Spring 1993	85		86.7	91.9	97.1		8.12	2.59	0.009
		41.2				22.5			
Summer 1993	120		105.9	112.6	119.2		−6.17	−2.18	0.029

PRB: percent relative bias = $100 \cdot (\hat{N} - N)/N$.
Z: standardized difference = $(\hat{N} - N)/SE(\hat{N})$.

References

Alonso, Robert S., et al. “Mark-Recapture and Mark-Resight Methods for Estimating Abundance with Remote Cameras: A Carnivore Case Study.” PLOS ONE, edited by Benjamin Lee Allen, vol. 10, no. 3, Mar. 2015, p. e0123032. DOI.org (Crossref), doi:10.1371/journal.pone.0123032.

Jessup, David. “Evolution of Ungulate Capture Techniques in California.” California Fish and Game, vol. 100, no. 3, 2014, pp. 491–526

Menkens, George E., and Stanley H. Anderson. “Estimation of Small-Mammal Population Size.” Ecology, vol. 69, no. 6, 1988, pp. 1952–59. JSTOR, doi:10.2307/1941172.

Vincent, Jean-Paul, et al. “Testing Density Estimators on a Fallow Deer Population of Known Size.” The Journal of Wildlife Management, vol. 60, no. 1, Jan. 1996, p. 18. DOI.org (Crossref), doi:10.2307/3802035.