ST 432 Kenneth H Pollock
Draft Notes January 23, 2009

NC Wildlife Commission 2007-2008 Hunter Survey Design and Analysis

Kenneth H. Pollock, Professor, and Zhi Wen, Graduate Student
Departments of Biology (and Statistics)
North Carolina State University
Box 7617 Raleigh NC 27695-7617

Introduction

I decided to use this survey that a graduate student and I are currently analyzing as an example of some of the topics covered in our class. The North Carolina Wildlife Commission ran a state-wide mail Survey of Hunting in 2008. Three of the key variables measured in the survey are Numbers of Hunters, Hunter Effort in Days, and Hunter Harvest for a wide variety of game species. The results of the survey will help inform managers of the species on how to adjust future hunting regulations.

Sampling Frame

The sampling frame used to draw the sample of hunters to contact consists of all of the hunter license files available to the Wildlife Commission. There are at least 16 types of licenses ranging from lifetime licenses to 1-year licenses to short term 6-day licenses purchased by nonresidents of NC. The period covered by the survey was from March 1, 2007 to February 29, 2008. The computerized license frame consisted of N = 482,588 hunters.

Under coverage due to some people not buying a license does occur but its extent is unknown. Another problem is that many licenses can be purchased on any day of the year and run for a year. This means that there are hunters in the frame that only have small parts of the year when they could legally hunt even if they bought a long-term license. We did not adjust for this in the sampling design or in the analysis.

Sampling Design

A 2% simple random sample from the whole frame was taken with a potential sample of n = 9652. We considered using a stratified random sampling design where the strata were different license types, but decided it was not necessary. [Any analyses on individual license types can use the subpopulation theory (Thompson 2002, Chapter 5)]. This size sample was based on prior studies and resulted in reasonable estimates for common species such as deer.

Contact Method

Mail surveys are a very cost efficient method of obtaining survey information, but they often suffer from substantial non response. Non response may cause bias in the population estimates of harvest and effort because the more avid hunters are more likely to respond. Wildlife Commission staff mailed questionnaires (Appendix) to a random sample of 2% of the licensed hunters. The frame included all hunters who had a valid license during one or more days between the period between March 1, 2007 and February 29, 2008, regardless of when the license was purchased. A modified Tailored Design method (Dillman 2000) with 3 mailings was used to try and minimize non-response. With this approach we expected a total response rate over all three mailings of 60-65 % based on past surveys. The first full questionnaire was mailed at the end of March with a reminder postcard a few days later. Two follow up questionnaire mailings were sent to non-respondents four weeks and 9 weeks later.

Time Table for the Survey

January-March 2008 Planning the Design of the Survey in collaboration with Wildlife Commission staff. Wildlife Commission staff prepared the survey which included preparation of the questionnaire and computer files for sampling, etc. They drew the sample in March. I was working with my student to develop an analysis plan.
April-June 2008. The survey ran by the Wildlife Commission staff and the data files edited and checked. Due to the three mailings it took about 9 weeks to get responses.
July-December, 2008. We analyzed data and a draft report was presented to Wildlife Commission staff. They will use the results of the survey to manage the species. We are currently making changes to the revised report and adding copies of our programs.

Analysis Methods and Results

Data Imputation and Related Issues

Before we could begin the formal statistical estimation analyses we had to make various adjustments and also resolve some complex data imputation issues. The initial frame size was 482,588 licensed hunters and the initial sample size of potential hunters to be contacted was 9,652 hunters for a target of a 2% initial sample. We found that there was an observed rate of deceased and ineligible hunters of 0.0071 in our sample which when applied to our frame size gave an adjusted frame size of 479,138 potential hunters. Deducting the ineligibles, deceased and non-deliverables the sample size was 8385 and of these 4716 responded for an adjusted response rate of 56.37 % over the 3 mailings. This is 4% lower than in the previous survey in 2005-2006.

We imputed for item non-response using an adaptation of the cell mean imputation method (Lohr 1999 p.272). This involves replacing the missing value with the mean value over other respondents who responded to the same variable. For example, if a known squirrel hunter left days hunted blank we imputed the mean number of days hunted based on all the squirrel hunters who did respond to that variable. We also had to include as non-response imprecise response statements such “hunted all season” when an exact number of days were asked for. In addition we found there were various outliers in the data that had to be treated as non-response such as number of days hunted beyond season length or number killed so large as to clearly be an error. The number of imputations was quite small for a particular variable so we ignored this aspect when computing standard errors of our estimates.

Population Estimates

We used standard simple random sampling estimation equations for population totals and their standard errors (Thompson 2002 p.16). An unbiased estimate of τ, the population total is:

$$\hat{\tau} = N\bar{y} \ ,$$

with standard error

$$SE(\hat{\tau}) = \sqrt{\{N(N-n)s^2/n\}},$$

with *N* the adjusted frame size and *n* the sample size of respondents. The sample mean is $\bar{y}$ and the sample standard deviation is *s*. We did include the finite population correction factor in the standard errors although this is only a 2% sample so the reductions in the standard errors are very small over ignoring them.

We make the standard assumption that our sample of respondents was a random sample despite non-response. We investigate this further in the next section. The adjustments on frame size and sample size made to account for out of frame values (e.g. deceased hunters) were ignored in standard error calculations due to their very small number. We also take the standard approach of assuming that the use of cell-mean imputations for item non-response do not significantly inflate the standard errors. This is also very reasonable because the number of these imputations was very small (Lohr 1999 p.272).

Tables 1 present the estimates of population totals and their standard errors for numbers of hunters, numbers of hunter days, and number of animals harvested first for the State Overall, the 3 regions estimates are not shown here. Precision (i.e. standard error) of estimates are reasonable except for some species not frequently hunted (e.g. Bears, Hogs).

Investigation of Non Response and Non-Response Bias

Mail surveys are a very cost efficient method of obtaining survey information, but they often suffer from substantial non-response (Cochran 1978 p. 359, Lohr 1999 p. 255, Dillman 2000). Non-response may cause bias in the population estimates of harvest and effort because perhaps the more avid members of any group (here hunters) are perhaps more likely to respond (Cochran 1978, Dillman 2000). This mail survey followed the general tailored design method of Dillman (2000) to attempt to reduce non response. With this approach we expected a total response rate over all three mailings of 60-65 % based on past surveys. The response rate actually achieved was a bit lower at 56% over the 3 mailings.

We also considered various analysis approaches for looking at the non-response issue. First we considered the non-response rates for the different types of licenses. We found that the response rate of lifetime license holders was 68% while for 12 month license types it was only 49% and for the short term non-resident licenses it was a perhaps surprising 61%. We expected the non-residents would be likely to respond at a lower rate. We note, however, that the nonresident hunters with the 6 day licenses are only a tiny fraction of the sample.

We considered how the selected sample and the actual responding sample differed by age and sex. We found that there was some change in distribution which reflected that older men and women were responding at a higher rate than younger ones. For example, men over 60 moved from 21.22% of the initial sample of males to 26.54% of the male respondents, whereas women over 60 moved from 15.33% of the initial sample of females to 17.05% of the female respondents. However, we decided not to weight responses as there was not very much change in almost all categories.

To investigate non-response bias in more detail we considered mailing by mailing responses of 6 of the more common game species. We used the species Deer, Doves, Ducks, Quail, Rabbits and Squirrels. We examined the mean days hunted and the mean animals killed by mailing. We suspected there might be a pattern of more avid hunters responding sooner (i.e. in an earlier mailing) than less avid hunters. Our results show that for most species there is not a strong pattern in this direction over the 3 mailings for days hunted which is one measure of avidity.

Of course weak patterns in non-response over mailings does not guarantee that there is no non-response bias as the non-responding hunters could still be very different from the respondents. As it is impractical to try and follow-up non-respondents with a different survey method (Cochran 1978 p. 370) there is no way this can be determined directly and indirect methods like these are all we can do. We conclude that while there may be some non-response bias in our estimates we suspect it is reasonably small and decided as in past surveys not to attempt to adjust for non-response bias. We return to this point when we discuss future survey design and analysis.

Suggestions for Future Survey Design, Data Imputation, and Estimation

The response rate in this survey was 56% which is not as high as that reported in past surveys. If this drop continues in the future then non-response bias is likely to become an even more serious issue.

The precision of the estimates in this survey appear to be adequate except on some of the very rare species like hogs suggesting that sample size does not need to be increased. Using an overall license file survey is not optimal for sampling rare species but this is not a new problem. The only solution would be to have special permits for these species so that a smaller specialist frame was available. Unfortunately this is very unlikely to be practical for the agency due to the high cost.

In future surveys both the license type and the number of days of validity of any license could be considered as stratification variables. There are pluses and minuses of such an approach which would need to be discussed and considered in detail. We are happy to do this with agency staff if they so desire.

We also are concerned about the need to change the question on number of days hunted. In the current survey there imprecise answers on number of days hunted where some respondents said that they hunted all season. We did not really know how to handle long seasons for this variable when hunters stated their response this way. The mean imputation method may underestimate their hunting activity but using the season length seemed a gross over estimate because that would imply they hunted coyotes every day of a 312 day season for example. Perhaps a better alternative would be to ask them to answer in large blocks of days from 1 up to the season length as this would be more likely accurate in their memories of the past year. It is practically impossible for a hunter to remember exactly how many days they hunted if the seasons are long and thus responses are likely to have serious measurement errors even if they did give a number rather than a vague response like hunted all season.

References

Dillman, D. A. (2000). Mail and Internet Surveys: The Total Design Method. John Wiley and Sons, New York, New York, USA.

Thompson, S. K. (2002). Sampling. Second Edition. John Wiley and Sons, New York, New York, USA.

Table 1: Estimated numbers of hunters, total days, and total harvest for game species, with standard errors (SE), in North Carolina 03/01/2007-02/29/2008, based on a mail survey of licensed hunters in North Carolina in Spring 2008. Overall State Estimates.

Species	Hunters	SE	Days	SE	Harvest	SE
Bear	18,102	1306	132,031	14,882	3,148	613
Geese	16,888	1,264	80,193	9,604	70,890	11,612
Coyotes	23,967	1,487	200,003	22,829	36,144	6,039
Bucks	--	--	--	--	149,877	4,862
Does	--	--	--	--	163,965	7,143
Deer	239,366	2,266	3,662,832	79,822	313,842	10,168
Doves	100,216	2,595	331,994	15,964	1,503,095	80,314
Ducks	41,664	1,898	567,219	221,169	469,227	43,565
Foxes	6,068	772	40,712	7,859	6,472	1,468
Grouse	8,393	904	43,383	7,698	8,745	2,709
Hogs	8,292	899	47,389	8,568	13,970	7,727
Quail	27,608	1,586	147,159	16,322	228,964	31,150
Rabbits	62,395	2,229	418,716	25,381	382,407	29,293
Raccoons	19,012	1,336	264,660	35,161	92,053	14,236
Squirrels	69,878	2,322	431,268	26,989	482,206	33,270
Turkeys	72,609	2,353	400,489	17,789	28,161	3,597
Woodcock	2,832	530	17,863	4,780	4,871	1,500