AN ADJUSTMENT FOR NON-RESPONSE BIAS IN A MAIL-OUT GAME HARVEST SURVEY

ALBERT W. GREEN, Texas Parks and Wildlife Department, Austin, TX 78744
JOHN C. BARRON, Texas Parks and Wildlife Department, Austin, TX 78744

Abstract: The Texas Parks and Wildlife Department conducted 2 harvest surveys by mail during 1976, 1 sampled currently licensed hunters and the other sampled hunters that were licensed 1 year previously. Both surveys requested information concerning hunting activity during the current season. A portion of the currently licensed hunters who did not return questionnaires were contacted by telephone to adjust for non-response bias. The survey sampling out-of-date hunters was not adjusted for non-respone bias. The harvest and related hunting statistics estimated from regression models developed from the result of these 2 surveys are comparable to statistics estimated by sampling currently licensed hunters. This technique also provides for the early availability of harvest statistics (6 weeks after the last day of the season) and reduces the expense of conducting the survey by eliminating the telephoning requirement to adjust for non-response.

Mail surveys are an important means of obtaining information in many fields. Wildlife managers are increasing their use of this technique to secure hunter and harvest estimates in lieu of more traditional (and costly) personal contacts such as check stations. A major source of error which must always be considered in mail surveys is the problem of non-response bias. It is well known that people who do not answer surveys and questionnaires tend to form statistically different populations from people who do answer. Cochran (1967) has discussed this phenomenon and the analysis of non-response. Several workers have studied the specific non-response bias associated with game harvest surveys (Sen 1972, Filion 1974).

Since 1972 the Texas Parks and Wildlife Department has conducted a mail survey of hunting license holders to determine the harvest of white-tailed deer (*Odocoileus virginianus*), mule deer (*O. hemionus*) and wild turkey (*Meleagris gallopava*). Clark and Boydston (1973) described the design which was patterned after a survey developed by Eberhardt and Murrey (1960). The Texas survey used only current license holders in hopes of increasing response and avoiding other sampling problems. In spite of this precaution a substantial proportion did not answer the mail surveys and this non-response stratum was sampled by telephone. The result obtained by contacting non-respondents was used to adjust the harvest and hunter estimates.

We found, as did Hawn and Ryel (1969) that this type of survey is slow and expensive. Though highly useful, the harvest statistics took from 4 to 6 months to obtain and cost in excess of $30,000 a year. In an effort to determine if these costs could be reduced without compromising the integrity of the survey, an experiment was made in 1976 in which a portion of the persons sent questionnaires were license holders from the previous year.

After having conducted a big game harvest survey for 3 years in Texas, we found that a large proportion of hunters hunting in any given year also hunted the year before. Thus it should be possible to sample a year old license list and obtain information concerning a current hunting season. Sampling from a year old license list would permit an earlier mail-out since labels and questionnaires could be prepared before the current hunting season was over. The effectiveness of such a procedure was evaluated using regression techniques. The objective was to develop prediction models for harvest and hunter estimates for the current hunting season using estimates obtained from a year old license list.

We gladly acknowledge G.A. Boydston of the Texas Parks and Wildlife Department as the driving force behind this project. It was his continued insistence that the survey be made more responsive and his continued support of the project that made the development of this technique possible. We would also like to thank our colleague D.M. McCarty for assistance with computer programming.

METHODS

The data for the regression models were collected by conducting 2 separate mail-out surveys during the 1975 hunting season. The sampling frame for the primary survey was a current license list (licenses sold during the period 1 September 1975 through 4 January 1976). The secondary survey sampling frame was a year old license list (licenses sold from 1 September 1974 through 4 January 1975). Both surveys requested information on the hunter's activities during the current hunting season.

Three statistics were computed from the primary survey: total harvest, number of hunters, and number of hunter days. These estimates were adjusted for non-response bias using data collected during a telephone follow-up (Clark and Boydston 1973). This report describes only the white-tailed deer results, but all 3 species in the survey were found to conform to these models. The Texas Parks and Wildlife Department has divided the state into 48 big game data reporting units (DRU's). Each DRU is composed of 2-14 counties which wildlife managers consider to have similar big game populations. Harvest estimates made for each of these DRU's were used to develop the regression models.

Linear regression models of the following type were computed from the survey results:

$$Y = a + b X.$$

Where a and b are constants estimated from the data and Y is the primary survey estimate (harvest, hunters, or hunter-days) which has been adjusted for non-response bias. X is a mean per respondent from the secondary survey. The mean was calculated as:

$$\overline{X} = X' (n-n').$$

Where X' is the number of hunters, hunter days or harvest reported for a specific DRU, n is the statewide total number of persons that responded to the secondary survey, and n' is the number of respondents in the secondary survey who did not report where they were hunting. The regressions were calculated using estimates (harvest, hunters, or hunter-days) from the primary survey as the dependent variable and estimates from the secondary survey as the independent variable.

RESULTS

In the primary survey 37,282 questionnaires were mailed: 2,499 of these were returned undelivered. After 2 mailings 20,157 persons had responded, leaving 14,626 (42.0%) as non-respondents. To correct for the non-response bias, 463 randomly chosen persons from this group were contacted by telephone. The secondary survey consisted of 8,044 mailouts with 888 returned as undeliverable, almost double the primary rate. Two mailings resulted in 3,848 respondents and 3,315 (46.3%) non-respondents. Only 1 area (DRU-46) failed to have any data reported by the surveys, thus reducing the potential number of data points from 48 to 47 for each regression model.

White-tailed deer harvest estimates (Y_K) ranged from 0 in DRU-17 to 4.57×10^4 in DRU-26. The corresponding mean kill per respondent (X) estimated from the secondary survey for each area was 3.00×10^{-4} and 6.47×10^{-2} respectively. The equation of this line was

$$Y_K = 8.129 + 7.303 \times 10^5 X_K$$

which accounted for over 95 % of the total variation ($r = 0.976$, $P < 0.01$).

Hunter estimates (Y_H) ranged from 78 in DRU-17 to 4.96×10^4 in DRU-25 and has corresponding X values of 3.00×10^{-4} and 0.048. Again the regression accounted for over

95% of the total variation with

$$Y_H = 787.7 + 9.46 \times 10^5 X_H$$

and r = 0.979 for n = 47.

Similar results were obtained for hunter-day estimates (Y_D). The regression was

$$Y_D = 5{,}555 + 8.61 \times 10^5 X_D$$

with a correlation coefficient of r = 0.963 for n = 47.

Prediction limits (95%) for the white-tailed deer harvest wre ± 4,650 deer at X = 0.009. Prediction limits for regression estimates for white-tailed deer hunters were ± 5,300 at X = 0.014. Limits for hunter-days were ± 42,400 at X = 0.086.

DISCUSSION

This study shows that a year old license list can be used to obtain estimates of current hunting activity that are highly correlated with estimates obtained from a current license list. Since the constants in the regressions were estimated using estimates adjusted for non-response bias, the regression estimates are similarly adjusted and it is not necessary to conduct a telephone follow-up survey. This results in a savings of time, salaries and a telephone bill.

Regression estimates are available from 6 to 7 weeks after the last day of the hunting season. The initial mailing of survey forms can begin on the last day of the hunting season, and a second mailing can be made at the end of 2 1/2 weeks. All responses received in the 3-week period after the second mailing can be edited and coded for computer processing as time permits during the fourth and fifth weeks and those mailers received during the fourth and fifth weeks can be coded and edited during the sixth week. Estimates can be prepared during the seventh week, soon enough to make harvest recommendations for the next season.

The precision of the estimates made by the models presented here could be improved upon greatly by enlarging the sample size of the secondary survey. Regression analysis assumes that the independent variable is measured without error. Since this is not the case in this situation, the error involved in estimating the independent variable should be as small as possible and this can only be done by increasing the sample size of the secondary survey. Another step that could be taken to increase the precision of these estimates is to repeat the experiment for more than 1 year.

These models should not be considered perfected at this time. The simple models presented are sensitive to temporal or geographic variations in hunting success or a redistribution of hunting pressure between DRU's but they are not sensitive to annual changes in total hunting pressure. The mean per respondent (X_k, X_H or X_D) for a given DRU will vary with changes in hunting success or with a redistribution of hunters between the DRU's. Unfortunately the means per respondent do not vary with changes in total hunting pressure. Therefore, the simple regression models that were presented are probably a special case of a more general multiple regression model

$$Y = a + bX + cZ$$

where Y and X are the same as defined previously and Z is the total license sales (i.e. potential hunters) for a given year. Constants estimated by the regression are a, b, and c (a and b are not necessarily the same as before). It was not possible to analyze the data using this model because the total number of licenses sold in Texas has not varied over 5% for the last 5 years. It is probably not possible to estimate c with such a small amount of variation in license sales. And as long as license sales remains essentially constant the use of the simpler models seems justified.

LITERATURE CITED

Clark, T.L., and G.A. Boydston. 1973. 1972-73 Texas deer-turkey harvest survey. Texas Parks and Wildl. Dept. Special Admin. Rep. 55 pp.

Cochran, D.G. 1967. Sampling techniques. 2nd ed. John Wiley & Sons, Inc. N.Y. 413 pp.

Eberhardt, L., and R.M. Murray. 1960. Estimating the kill of game animals by licensed hunters. Proc. Annual Meeting. Am. Stat. Assoc. 120:182-188.

Filion, F.L. 1974. Estimating bias due to non-response in harvest surveys. Can. Wildl. Serv. Manuscript Rep. No. 5. 57 pp.

Hawn, L.J., and L.A. Ryel. 1969. Michigan deer harvest estimates: sample surveys versus a complete count. J. Wildl. Manage. 33(4):871-880.

Sen, A.R. 1972. Some nonsampling errors in the Canadian waterfowl mail survey. J. Wildl. Manage. 36(3):951-954.