# DEVELOPMENT OF EXPLANATORY AND PREDICTIVE MODELS FOR HUNTING AND FISHING LICENSE SALES AND REVENUE TRENDS IN NEW YORK

PROGRESS REPORT NO. 2

by

Tommy L. Brown and Nancy A. Connelly

October 1986

Series No. 86-4

Human Dimensions Research Unit
Department of Natural Resources
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, N. Y.

# DEVELOPMENT OF EXPLANATORY AND PREDICTIVE MODELS FOR HUNTING AND FISHING LICENSE SALES AND REVENUE TRENDS IN NEW YORK

PROGRESS REPORT NO. 2

by

Tommy L. Brown and Nancy A. Connelly

October 1986

PROGRESS REPORT

STATE: NEW YORK

PROJECT NO.: W-146-R:12

PROJECT TITLE: Public Attitudes Toward Wildlife and Its Accessibility

STUDY NUMBER AND TITLE: I - Deriving Social Indices of Public Attitudes Toward Wildlife Populations and Their Use

JOB NUMBER AND TITLE: I-11 - Monitoring Trends in Human Population Characteristics of Importance to Wildlife Program Planning

JOB OBJECTIVES: To analyze the relationships between New York State population and user demographic trends and license sales performance.

To assist in wildlife program evaluation and planning by predicting trends in resource use and revenue levels associated with such use.

JOB DURATION: 1 July 1986-30 June 1989

The 1984-85 prediction errors in adjusted license sales were 7.1% for small game, 2.2% for big game, 5.8% for combination hunting-fishing, and 5.8% for fishing licenses. The stability of the models for small game, big game, and combination licenses was judged to be very good. Sufficient changes occurred in the coefficients of the fishing model, however, that although the 1984-85 error rate was reasonable, we are not confident that future models will continue to be at this level of accuracy. We continue to seek time series data that reflect fishing opportunities and quality for the State as a whole.

We have only begun analyses of the 1962-1983 combined county-level data base and the 1980 county-level cross sectional data base. Some additional independent or explanatory variables were available for these data bases. In several cases, population density was negatively associated with license sales. In the 1980 cross-sectional data base, the percent of the population over 5 years of age who lived in the same county as in 1975 was positively associated with big game license sales. Because of the high variability in the number of licenses sold across counties, the prediction errors of the first models reported on herein were quite large. We will attempt cluster analyses in the coming year to develop groupings of counties having similar characteristics that can be analyzed together to reduce the prediction error. We will also determine whether regional models of acceptable accuracy can be developed.

In summary, we recommend continuing this effort in 1986-87. In particular, we feel that the Human Dimensions Research Unit should pursue the following:

1. Seek new demographic data sources from state agencies (e.g., Labor, Education, Commerce).
2. Continue to seek resource data from Bureau of Fisheries (there has been no problem in the cooperation of Fisheries staff to specific

inquiries but we probably need to do some joint brainstorming of potential measures of fishing quality statewide).

3. As new variables are found that are statistically significant in one data base, strive to get them into other data bases. As of now, for example, we have not looked at population density in our statewide longitudinal model.

4. Pursue work to reduce the variance of license sales per capita in the two county-level models. Perform a cluster analysis of counties from the 1980 data base for each major license type, using license sales per capita and a few key demographic variables. Choose meaningful groupings of counties and run regression models on each group in an attempt to increase the accuracy of estimates and further define influential demographic variables.

TABLE OF CONTENTS

# INTRODUCTION

This is the second of what is anticipated to be a series of reports over time dealing with factors affecting hunting and fishing license sales in New York, and the revenue implications of those factors. This report references heavily the August, 1985 Progress Report No. 1 of similar title; that report should be referred to in conjunction with this report.

The purpose of this series of reports is to develop a better understanding of the resource and demographic forces that influence hunting and fishing license sales in New York and the resulting revenue generated to the Division of Fish and Wildlife. This report utilizes three data sets:

1. A state-level longitudinal data base from 1962-1984. Models from this data base were reported on in the 1985 progress report. At this time, an additional year's license sales data are available, so Part I of this report updates that material. It does so by inserting 1984-85 license sales data and demographic estimates in the models developed previously, evaluating the amount of error in model projections for 1984-85, and ascertaining the degree to which the insertion of 1984-85 data changes the variables and coefficients in each model.

2. A combined county-level data base for all counties for the years 1962-1983. Depending on data availability, 903 to 1,263 cases or observations comprise each model.

3. A 1980 county-level cross-sectional data base. Because 1980 was a major census year, more demographic variables are available than in other years to examine as possible influences upon license sales.

Each of these data sets uses similar dependent variables: resident small game, big game, combination hunting-fishing, and fishing licenses. However, the independent variables differ according to availability, particularly with respect to demographic variables.

It should be noted carefully at this point that each of the 3 data bases examines license sales from a different perspective. As a result, even if the same measures of the same independent variables were available for each data set, we would not expect nearly identical models (with the same independent variables and similar coefficients) to be developed for each type of license sales analyzed. This will become clearer as the perspective for each model is described below.

The longitudinal, state-level data base allows the development of models that show how total statewide sales of a particular type have been influenced over more than 2 decades by changes in demographic factors, the cost of the license, and to the extent that we have measures, resource (supply and access) factors. These models do not capture time in any direct way (i.e., a 1962 observation has the same weight as a 1985 observation), but the residuals (the prediction errors) are examined to ascertain that they are not temporally correlated. The state-level data base has only 23 or 24 observations. Performing regressions with up to a dozen independent variables poses no theoretical problems, but it is more likely to find chance correlations with high r squares with a limited amount of data. Because we have only educated guesses of all of the relevant independent variables, those with high correlations must be examined very carefully to correctly choose the actual causal variables, or variables that are good indicators of causal variables when good measures of the causal variables are not available.

The combined longitudinal county-level data base examines those factors associated with changes in a given type of license sales both between counties and over time. Thus, it is both longitudinal and cross sectional. Its advantages are that it provides a large data base from which one can feel more confident that correlations are meaningful and not just chance events. Because it is longitudinal, the effects of changes in the license fee can be examined. The primary disadvantage is that New York's counties form such diverse demographic and resource continua (e.g., from Hamilton to New York County) that it is unlikely that variables can be found to measure the vast proportion of the variance in license sales. Some New York City counties will have lower amounts of participation, and some upstate counties such as Erie will have higher levels of participation than we can ever explain with the variables available for the analysis. To some degree, for the purposes of prediction, dummy variables (dichotomous variables coded "1" if a characteristic is present and "0" if it is not) of clusters of such counties can be formed. This procedure increases the fit of the model (i.e., r square and standard error), which may improve prediction, but it does not increase our understanding of why these counties are outliers with respect to license sales.

The 1980 cross-sectional data set provides a means of examining factors related to differences in the sale of licenses among counties in a fairly recent year. Models from this data set should provide some insight as to why license sales per capita are considerably higher in some counties than in others. One would expect significant independent variables found in models of this data base to also be acting over time, but this is not necessarily the case. If it is the case, a given independent variable may still have a somewhat different relationship to license sales over time than between counties at a given point in time. Finally, since everyone pays the same fee for the same license at a point in time, the cross-sectional model can not examine the effects of changes in license fees, and therefore few, if any, revenue implications can be drawn from models resulting from this data base.

Thus, each of the 3 data bases has different strengths and limitations as to insights that models derived from each can provide about past and future license sales and revenue implications. Because the 3 types of models yield 3 kinds of insights for a given type of license sales, it is important to attempt some synthesis of the findings. This is done sequentially. Independent variables entering the county-level models for small game licenses, for example, are compared with their counterparts in the statewide model.

## PART I: UPDATE OF STATEWIDE DATA BASE

An August, 1985 progress report "Development of Explanatory and Predictive Models for Hunting and Fishing License Sales and Revenue Trends in New York" by Brown presented models for resident small game, big game, combination hunting-fishing, and fishing licenses. To permit a consistent data base over time, certain resident license types were combined to form the 4 groupings used in that report and continued in this report:

1. Small game. Small game plus small game/big game combination licenses.
2. Big game. Big game plus small game/big game combination plus sportsman's licenses.
3. Combination hunting-fishing. Hunting-fishing plus sportsman's licenses.
4. Fishing. Regular fishing plus 3-day license.

Once combined as indicated above these 4 types of licenses are referred to as adjusted small game, big game, combination hunting-fishing, and fishing licenses.

Refer to the previous report for a list of available variables and for both considerations and precautions in interpreting multiple regression models (p. 5-9). This update evaluates for each of the 4 license types the degree of error in the previous model when the most recent year's data are inserted and the model predicts 1984-85 license sales. It then examines the degree to which another year's data changes the predictive model, and points out any implications of changes or lack thereof.

Small Game License Sales

For the period from 1961-62 through 1983-84, the best explanatory model for adjusted small game license sales, with standard errors in parentheses, was:

$$\begin{aligned} SGLS = & \underset{(194,763)}{-113,064} + \underset{(41.46)}{166.6}\ P45\text{-}64 - \underset{(22.87)}{83.17}\ NAg + \underset{(21.67)}{99.46}\ REAL \\ & - \underset{(4,020)}{8,046}\ \$\$ - \underset{(48.10)}{75.07}\ IM + \underset{(9,392)}{15,284}\ CL \end{aligned}$$

Where:

| | | |
|---|---|---|
| SGLS | = | Number of adjusted small game licenses sold; |
| P45-64 | = | 45-64 aged NY population (thousands); |
| NAg | = | Total nonagricultural employment in New York (thousands); |
| REAL | = | NY per capita income, adjusted for inflation; |
| $$ | = | License cost; |
| IM | = | Miles of interstate highway open in NY; |
| CL | = | Dummy variable indicating whether or not the combination small game/big game license was sold. |

This model had an adjusted $r^2$ of .957, standard deviation of 9,171 licenses, and a mean error of 6,211 licenses over the 22-year period, ranging from 14,964 in 1969 (a 5.8% error) to 799 in 1977 (a 0.3% error). The error pattern was not time-correlated.

If 1984-85 sales are projected from this model, the model predicts sales of 142,837 licenses sold compared to actual sales of 153,699. The model underpredicted license sales by 10,862, or 7.1%. This underprediction is primarily due to the drop in the 45-64 aged population and the increase in nonagricultural employment in 1984-85. Still, this error is within the range of previous years' predictions.

If the small game license sale data base is updated to include the 1984-85 license sales, costs, and estimates of the demographic variables[1], a new small game model can be developed. The best new model uses the same variables as the previous model to predict small game license sales, but the coefficients of those variables have changed somewhat. Standard errors are in parentheses under each coefficient:

$$\begin{aligned} SGLS = {} & \underset{(172{,}853)}{-226{,}810} + \underset{(39.6)}{182.6}\ P45\text{-}64 - \underset{(18.9)}{70.5}\ NAg + \underset{(20.2)}{89.2}\ REAL \\ & \underset{(3507)}{-5464}\ \$\$ - \underset{(46.9)}{82.1}\ IM + \underset{(8690)}{20{,}139}\ CL \end{aligned}$$

This model has an adjusted $r^2$ of .965 and a standard deviation of 9,155 licenses. The explanatory power of this model is slightly better than that of the previous model. The changes in the values of the coefficients are examined below for their implications.

The new model has higher coefficients, and is therefore more sensitive to the 45-64 age population and the dummy variable indicating whether or not the combination small game/big game license was sold. The population in this age bracket is projected to increase at least through 1995, so the higher coefficient should exert a stronger positive influence on the number of small game licenses sold than predicted by the previous model. Similarly, the larger coefficient of the small game/big game license dummy variable will exert a stronger positive influence on licenses sold. A decrease in the absolute value of the coefficient associated with the cost of the small game licenses implies that small game license sales will be even less sensitive to a license fee increase than previously predicted.

---

[1]Only population projections are available. The authors have projected other demographic variables, based on recent trends and other relevant information.

Thus, projections for small game license sales and revenues, including the possibility of a license fee increase, seem relatively positive from this model. The authors still share concern, however, that declining hunter training course registrations, not covered by the model, will exert an additional dampening effect on both small game and big game license sales.

The projections of the independent variables shown in Table 1 were substituted into the new small game model, varying the license fee by $1.00 increments to examine projected effects on revenue. The model estimates a loss of 5,464 adjusted small game licenses sold for each dollar that the fee is increased. Estimated sales and revenue generated for a schedule of possible fees is shown in Table 2.

Table 1. Projected Independent Variable Data Affecting Small Game License Sales, Excluding License Cost.[a]

| Variable | 1987 | 1990 |
|---|---|---|
| Pop. 45-64 | 3,646,000 | 3,655,000 |
| Nonag. Employment | 7,650,000 | 7,750,000 |
| Real Per Capita Income (1967 Dollars) | 4,650 | 4,800 |
| Interstate Miles | 1,525 | 1,535 |
| Presence of Small Game/ Big Game License | Yes | Yes |

[a]Only population projections are available. The authors have projected other variables, based on recent trends and other relevant information.

Table 2 shows that according to the new small game model, revenue would not be maximized in 1987 or 1990 at a fee of twice the current fee. A weighted $16.37 fee undoubtedly is not politically feasible, and would result in the loss of over 43,000 small game licenses, but according to the model, this fee would increase revenues by about $600,000. The inability to maximize revenues even at twice the current fee illustrates the finding that license sales are not highly sensitive to the amount of the license fee.

Table 2. Projected Small Game License Sales and Revenues for 1987 and 1990.

| Year | Weighted Fee | Licenses Sold | Revenue (000's) |
|---|---|---|---|
| 1987 | $ 8.37[a] | 163,337 | $1,367 |
| | 9.37 | 157,873 | 1,479 |
| | 10.37 | 152,409 | 1,580 |
| | 11.37 | 146,945 | 1,671 |
| | 12.37 | 141,481 | 1,750 |
| | 13.37 | 136,017 | 1,819 |
| | 14.37 | 130,553 | 1,876 |
| | 15.37 | 125,089 | 1,923 |
| | 16.37 | 119,625 | 1,958 |
| 1990 | $ 8.37 | 170,376 | $1,426 |
| | 9.37 | 164,912 | 1,545 |
| | 10.37 | 159,448 | 1,653 |
| | 11.37 | 153,984 | 1,751 |
| | 12.37 | 148,520 | 1,837 |
| | 13.37 | 143,056 | 1,913 |
| | 14.37 | 137,592 | 1,977 |
| | 15.37 | 132,128 | 2,031 |
| | 16.37 | 126,664 | 2,073 |

[a]1985 weighted small game license fee.

## Big Game License Sales

For the period from 1961-62 through 1983-84, the best explanatory model for adjusted big game license sales, with standard errors in parentheses was:

$$\begin{aligned} \text{BGLS} = & \underset{(168{,}216)}{-329{,}494} + \underset{(36.71)}{86.53}\ \text{P18-44} + \underset{(0.91)}{6.22}\ \text{BH-1} - \underset{(0.23)}{1.20}\ \text{P} \\ & \underset{(7{,}284)}{-18{,}586}\ \$\$ + \underset{(57.13)}{149.37}\ \text{IM} + \underset{(28{,}354)}{65{,}136}\ \text{CL} \end{aligned}$$

Where:

BGLS = Big game license sales;

P18-44 = NY 18 to 44 aged population;

BH-1 = Adult bucks harvested the previous year;

P = Number of deer management permits issued during the current year;

$$ = Weighted license fee;

IM = Miles of interstate highway open in NY;

CL = Dummy variable indicating whether a combination small game/ big game license was sold.

This model had an adjusted $r^2$ of .960, standard deviation of 17,122 licenses, and a mean error of 10,826 licenses over the 22 year period, ranging from 33,757 in 1971 (a 5.8% error) to 96 in 1974 (a 0.02% error). The error pattern was not time correlated, but the average error over the last 4 years (1981-1984), 3,838 licenses, was well below the mean error of 10,826 for the entire 22 years.

If 1984-85 sales are projected from this model, the model predicts sales of 672,412 licenses, compared to actual sales of 658,072. The model overpredicted license sales by 14,340, or by just 2.2%. This small overprediction is primarily associated with a more rapidly increasing population, aged 18-44, in the past two years than in previous years of the model.

If the data base is updated to include 1984-85 license sales, costs, and estimates of the demographic variables, a slightly different, updated model can be developed. The best new model uses the same variables as the previous model to predict license sales, and the coefficients of most variables are of very similar magnitude to the previous model. Standard errors are in parentheses under each coefficient:

$$\text{BGLS} = \underset{(154{,}258)}{-345{,}912} + \underset{(33.50)}{89.61}\ \text{P18-44} + \underset{(0.82)}{6.38}\ \text{BH-1} - \underset{(0.21)}{1.24}\ \text{P}$$

$$-21{,}512\ \$\$ + 151.77\ IM + 64{,}477\ CL$$
$$(6{,}651)\qquad (53.80)\qquad (27{,}061)$$

This model has an adjusted $r^2$ of .963 and a standard deviation of 16,493 licenses. The explanatory power of this model is slightly better than that of the previous model. The only change in the value of a coefficient worth further examination is that of the weighted license fee.

In the new model the absolute value of the coefficient of the license fee variable has increased by about 16%. This increase implies that big game license sales will be somewhat more sensitive to license fee increases than previously predicted. An increase of $1.00 in the license fee would result in a loss of 21,512 big game licenses sold, compared to 18,586 in the previous model.

Because we are unable to project sales and revenues for 1987 and 1990 without estimates of deer permits issued and buck harvest, we will examine the revenue impacts of increasing the big game license fee for the last year for which data are available, 1984-85. Table 3 shows that the big game license revenue is maximized at about $20.28, slightly lower than the previous model that maximized revenue at a fee of $22.19. However, this maximizing point is far above the current, weighted fee of $8.28. Thus, big game license sales for the vast majority of hunters is not very price sensitive at its current level. Unless deer populations drop severely, there is solid evidence that a license of any politically acceptable amount would increase big game revenues.

Combination Small Game/Fishing License Sales

For the years 1961-62 through 1983-84, the best explanatory model for adjusted combination small game/fishing license sales was:

Table 3. Projected Big Game License Sales and Revenues in 1985, for a Schedule of Possible Fees.

| Weighted Big Game Fee | Projected License Sales | Projected Revenues (000's) |
|---|---|---|
| $ 8.28 | 661,359 | 5,476 |
| 10.28 | 618,335 | 6,356 |
| 12.28 | 575,311 | 7,065 |
| 14.28 | 532,287 | 7,601 |
| 16.28 | 489,263 | 7,965 |
| 18.28 | 446,239 | 8,157 |
| 20.28 | 403,215 | 8,177 |
| 21.28 | 381,703 | 8,123 |

$$
\begin{aligned}
SGFC = {} & \underset{(165{,}822)}{957{,}918} - \underset{(39.59)}{194.56}\ P45\text{-}64 + \underset{(14.43)}{91.67}\ REAL - \underset{(1{,}866)}{12{,}436}\$\$ \\
& - \underset{(8{,}949)}{32{,}324}\ CONT + \underset{(10{,}685)}{43{,}749}\ SPORT - \underset{(47.83)}{172.44}\ P14\text{-}17
\end{aligned}
$$

Where:

SGFC = Small game/fishing combination license sales;

P45-64 = NY population 45-64 years of age;

REAL = Per capita income, adjusted for inflation;

CONT = Dummy variable indicating presence of contaminants in Lake Ontario waters;

SPORT = Dummy variable indicating whether or not sportsman's license was sold that year;

P14-17 = NY population 14-17 years of age.

This model had an $r^2$ of .926 when adjusted for degrees of freedom, and a standard deviation about the regression line of 8,735 licenses. Errors in license sales average 6,222 licenses, and the errors are not time correlated.

When the relevant 1985 data are added to the data base, the model predicts 290,373 adjusted combination licenses sold, compared to actual sales of 274,262. This overprediction of 16,111 licenses represents a 5.8% error.

An updated model for combination licenses which uses the same variables, with standard errors in parentheses, is:

$$
\begin{aligned}
\text{SGFC} &= \underset{(207{,}375)}{830{,}097} - \underset{(49.34)}{163.44}\ \text{P45-64} + \underset{(18.68)}{87.07}\ \text{REAL} - \underset{(2{,}832)}{12{,}365}\ \$\$ \\
&\quad - \underset{(12{,}084)}{15{,}299}\ \text{CONT} + \underset{(13{,}414)}{38{,}469}\ \text{SPORT} - \underset{(60.36)}{157.48}\ \text{P14-17}
\end{aligned}
$$

This model fits the data less precisely than the previous model fit data through 1984. This model has an adjusted $r^2$ of .879, and a standard error of the estimate of 11,202 licenses, compared to 8,735 for the previous model. The explanation for the less precise fit may lie with the contaminant dummy variable. With the inclusion of 1985 data, this variable attained a much larger standard error. The implication is that the contaminants, which up until 1985 were "costing" the sale of about 32,000 licenses, are no longer retarding combination license sales (note the current coefficiant of -15,299 and standard error of 12,084). For comparative purposes, because the contaminant variable may still be acting weakly and its omission from the model does not improve the fit of the model, we will use this as the best current model of combination license sales.

The coefficient of the new model associated with the license fee, -12,365, is approximately that of the previous model of -12,436. The data used to project sales and revenue to 1987 and 1990 are shown in Table 4.

Revenue estimates generated by the model are slightly less than for the previous model, but the structure by possible fees is very similar. Peak revenue in 1987 would be generated at a fee of \$19.81, and in 1990 at a fee of \$21.81 (Table 5).

Table 4. Projected Independent Variable Data Affecting Combination Small Game/ Fishing License Sales, Excluding License Cost.[a]

| Variable | 1987 | 1990 |
|---|---|---|
| Population 45-64 | 3,646,000 | 3,655,000 |
| Real Per Capita Income (1967 Dollars) | 4,650 | 4,800 |
| Presence of Lake Ontario Contaminants (Dummy) | Yes | Yes |
| Sale of Sportsman's License (Dummy) | Yes | Yes |
| Population 14-17 | 1,024,000 | 890,000 |

[a]Per Capita Income projections are not available but have been projected by the authors.

Table 5. Projected Combination Small Game/Fishing License Sales and Revenues for 1987 and 1990.

| Year | Weighted Fee | Licenses Sold | Revenue (000's) |
|---|---|---|---|
| 1987 | 15.81 | 305,490 | 4,830 |
| | 17.81 | 280,760 | 5,000 |
| | 19.81 | 256,030 | 5,072 |
| | 20.81 | 243,665 | 5,071 |
| | 21.81 | 231,300 | 5,045 |
| 1990 | 15.81 | 338,182 | 5,347 |
| | 17.81 | 313,452 | 5,583 |
| | 19.81 | 288,722 | 5,720 |
| | 21.81 | 263,992 | 5,758 |
| | 22.31 | 257,810 | 5,752 |
| | 22.81 | 251,628 | 5,740 |

As noted in the earlier report, combination license sales are very strongly correlated (positively) with the 18-44 age population. However, the high (negative) correlation between the 18-44 and 45-64 age groups prohibits the former variable from entering the model.

Fishing License Sales

Using data through 1984, the best explanatory model for adjusted resident fishing licenses was:

$$\underset{}{FLS} = \underset{(276,580)}{748,715} + \underset{(17.97)}{85.69}\ REAL + \underset{(14,132)}{60,745}\ USAL - \underset{(64.22)}{129.83}\ P45\text{-}64 - \underset{(6,300)}{12,079}\ \$\$$$

Where:

FLS = Adjusted resident fishing license sales;

REAL = Per capita income in 1967 dollars, adjusted for inflation;

USAL = Dummy variable representing the years 1973-75 when the Great Lakes salmonid fisheries opened, and before the discovery of contaminants.

P45-64 = NY 45-64 age population

$$ = license cost

This model has an $r^2$ of .730 when adjusted for degrees of freedom, and a standard deviation of predicted data around the regression line of 22,102 licenses.

This model predicted 1985 sales of 550,724 licenses, compared to actual sales of only 520,313. The overprojection of more than 30,000 licenses represents an error of 5.8%.

When 1984-85 data are inserted into the model, the coefficients and standard errors change as follows:

$$FLS = \underset{(299,932)}{623,914} + \underset{(21.05)}{72.62}\ REAL + \underset{(16,443)}{58,927}\ USAL - \underset{(6,899)}{7547}\ \$\$$$

This model fits the data less well, providing an adjusted $r^2$ of .65, and a standard deviation of predicted data about the regression line of 24,451 licenses.

The coefficient of the license sales variable has dropped from -12,079 in the previous model to -7,547 in the current model, and no longer approaches statistical significance at the .05 level (the current level is roughly .30). If the revised coefficient were accurate, it would indicate that anglers are even less sensitive to changes in license fees than calculated previously, when the projected 1987 revenue maximizing point was estimated to be at a license fee of $28.17.

The revised effects on revenue have not been calculated because we question the license cost coefficient; moreover, we are not satisfied with the overall model. We feel that we can draw general conclusions that fishing license sales are positively associated with real per capita income, and that license sales are not highly sensitive to the license fee in the area within a few dollars of current rates. But there are other opportunity factors (e.g., access, fishing quality), and perhaps some competing substitute variables that affect fishing that are not well understood, and are beyond any data base currently available to us.

We recommend an inquiry of fish managers regionally for any insights into changes in license sales regionally. If this seems productive (i.e., if managers feel they have plausible insights into generally declining sales in recent years), we should try to locate data bases that reflect these factors and incorporate them into our models.

# PART II:

## THE COMBINED LONGITUDINAL COUNTY-LEVEL DATA BASE

The combined longitudinal county-level data base provides a large data base (up to 1,263 county-by-year observations) from which one can feel more confident of the accuracy of identified predictive variables. However, the data base has at least two drawbacks. First, a limited number of demographic variables are available annually at the county level (e.g., total population is available, but not broken down by age categories or sex). Many of the demographic variables in these models are the same as ones found in the statewide or cross-sectional models. Second, the most recent data available to us for this data set are for 1983. We do not have county-level demographic projections for 1987 or 1990, and we do not feel confident in making such projections with the data available to us. Hence, the primary current use of this data set is to determine the extent to which variables that enter the statewide equation for a given type of license sales are also significant at the county level over time.

Because New York State counties are so diverse with respect to demographic and resource characteristics, it was difficult to develop statewide models with an acceptable level of predictive ability. Therefore, county-level models were developed for Upstate New York (i.e., counties north of Westchester and Rockland Counties). These models have average prediction errors of about 30% (statewide models ranged from 32 to 50%). Because of the lack of accuracy of statewide models from this data base, only the upstate models are reported in this section.

Small Game License Sales

The best explanatory model developed for upstate New York, small game license sales (with standard errors in parentheses) was:

$$\begin{array}{lllllll} SGLS & = & 925 & + 0.026\ POP & + 0.849\ REAL & - 6.43\ POPDEN & - 653.2\ \$\$ \\ & & (211) & (0.0006) & (0.046) & (0.51) & (73.7) \end{array}$$

Where:

SGLS = Number of adjusted small game licenses sold;

POP = Total county population;

REAL = Per capita income;

POPDEN = Population density measured in total population/square mile;

$$ = License cost.

This model has an adjusted $r^2$ of .873 and a standard error of 1,488 licenses or 32% of the mean number of 4,629 licenses sold per county per year for the 824 county-year cases.

Annual population by age groupings is not available at the county level, so total population was used. Total population was positively correlated with small game license sales (SGLS) at the county level. The 45 to 64 year group had the highest correlation (positive) in the statewide model. A population density variable (population per square mile) was established at the county level, and was negatively correlated with SGLS.

The proportion of people employed in agriculture by county was negatively correlated with SGLS at the county level, whereas total nonagricultural employment was negatively correlated with SGLS in the statewide model. At the county level, the counties with the highest proportions in farm employment are the most rural counties in the state with the lowest populations to draw small game hunters from. Farm employment is less than 1.5% of total employment statewide. The proportion of people employed in agriculture was negatively correlated with SGLS in both data sets. Comparable measures of these two employment variables were not available for both data sets for all years in the model. We will continue to search for the missing data.

Per capita income was positively correlated with SGLS for both the statewide and the county model. Similarly, the license cost variable had a significant negative correlation for both models. The negative coefficient of -653.2 indicates that for the average county over time, an increase of $1.00 in the license fee has been associated with the loss of 653 licenses, controlling for other factors statistically. The magnitude of this coefficient is somewhat suspect because the past two license fee increases have resulted in extremely minor changes in license sales. Regardless of the validity of this particular model (which can be ascertained only with improved data), one would expect the coefficient of the statewide model to be more accurate for statewide estimates than the expansion of a countywide model.

Big Game License Sales

The best explanatory model developed for upstate New York was:

$$\begin{aligned} \text{BGLS} = {} & \underset{(501)}{5{,}151} + \underset{(0.001)}{0.038}\ \text{POP} + \underset{(0.09)}{1.06}\ \text{BH-1} - \underset{(105)}{282.5}\ \$\$ \\ & - \underset{(20.7)}{102.5}\ \text{PAg} + \underset{(410)}{2{,}377}\ \text{UPSMSA} - \underset{(0.87)}{6.6}\ \text{POPDEN} \end{aligned}$$

Where:

BGLS = Number of adjusted big game licenses sold;

POP = Total county population;

BH-1 = Adult bucks harvested the previous year;

$$ = License cost;

PAg = Percent employed in agriculture;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more;

POPDEN = Population density measured in total population/square mile.

This model has an adjusted $r^2$ of .870 and a standard error of 2,435 licenses, or 29% of the mean of 8,361 licenses sold per county per year for the 823 cases.

Similar to the small game model, population has a positive influence and population density a negative influence on license sales. The number of adult bucks harvested in a county in the previous year has a positive effect on license sales in the current year. The cost of the license had the expected negative effect. For a $1.00 fee increase, an average of 282 licenses were lost per county, holding other variables constant.

Comparing the variables in the countywide and statewide big game models reveals that population, the previous year's buck harvest, and license cost are significant variables in both models. Because human population variables for which data are now available differ, direct comparisons can not be made. The statewide model uses the 18-44 population group. The countywide model uses total population, population density, and an SMSA dummy variable. Other significant statewide variables, deer management permits and interstate road miles, were not available at the county level. The dummy variable indicating sale of small game/big game combination licenses was not used in the countywide model because the most recent year's data for that model are from 1983. That variable will be incorporated as the data base is updated.

Combination Small Game/Fishing License Sales

The best explanatory model developed for upstate New York was:

$$\underset{}{SGFC} = \underset{(114)}{1{,}682} - \underset{(0.320)}{1.956}\ POPDEN - \underset{(24.1)}{213.6}\ \$\$ + \underset{(96)}{1{,}038}\ GLC$$

$$+ \underset{(161)}{1{,}977}\ UPSMSA + \underset{(0.028)}{0.359}\ REAL + \underset{(0.0004)}{0.018}\ POP$$

Where:

SGFC = Number of adjusted small game/fishing combination licenses sold;

POPDEN = Population density measured in total population/square mile;

$$ = License cost;

GLC = Dummy variable indicating counties bordering the Great Lakes;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more;

REAL = Per capita income;

POP = Total county population.

This model has an adjusted $r^2$ of .924 and a standard error of 1,065 licenses. The mean number of licenses sold for the 1,103 county-year cases was 4,131, which gives a mean error of about 26%.

Similar population, income, population density variables are present in this model as in models previously discussed. The license cost has a negative influence averaging 214 licenses per county per dollar of fee increase.

A dummy variable indicating the 9 counties along the Great Lakes is positively associated with license sales and accounts for an average increase of 1,038 licenses sold per county. This variable suggests a region along the Great Lakes with similar access and supply for fishing. A regional model developed for this area, which is a better predictor of license sales for the region, is:

$$\begin{aligned} GLSGFC = {} & \underset{(441)}{2285} + \underset{(0.001)}{0.010}\ POP + \underset{(0.099)}{0.643}\ REAL + \underset{(1075.5)}{3981.7}\ UPSMSA \\ & - \underset{(89.9)}{323.5}\ \$\$ + \underset{(1.19)}{2.79}\ POPDEN \end{aligned}$$

Where:

GLSGFC = Number of adjusted small game/fishing combination licenses sold in Great Lakes counties;

POP = Total county population;

REAL = Per capita income;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more;

$$ = License cost;

POPDEN = Population density measured in total population/square mile.

This model has an adjusted $r^2$ of .947 and a standard error of 1,545 licenses. The mean number of licenses sold for this sample of 187 cases was 7,875.

The variables are the same as in the upstate model and have similar coefficients, except population density, which is positively correlated with license sales for Great Lakes counties. This model more accurately reflects the situation along the Great Lakes where there are large numbers of licenses sold in the suburban areas of Erie and Monroe counties.

## Fishing License Sales

The best explanatory model developed for fishing license sales in upstate New York was:

$$FLS = \underset{(292)}{2{,}497} + \underset{(0.0008)}{0.034\ POP} + \underset{(261)}{2{,}473\ GLC} + \underset{(437)}{7{,}798\ UPSMSA}$$

$$+ \underset{(0.07)}{0.33\ REAL} - \underset{(111.5)}{172.6\ \$\$}$$

Where:

FLS = Number of adjusted fishing licenses sold;

POP = Total county population;

GLG = Dummy variable indicating counties bordering the Great Lakes;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more;

REAL = Per capita income;

$$ = License cost.

This model has an adjusted $r^2$ of .896 and a standard error of 2,896 licenses. The mean number of licenses sold for a sample of 1,103 cases was 8,933, which gives a mean error of about 32.5%.

In this model the license fee was negatively correlated with license sales and has been left in the model, but it was not statistically significant at the 0.05 level. As with the combination small game/fishing license sales, the Great Lakes dummy variable has a positive effect on license sales. A regional model for the Great Lakes is also a better predictor of fishing license sales in the Great Lake counties. The model developed for fishing license sales in the Great Lakes region is:

$$GLFLS = \underset{(1{,}162)}{5{,}213} - \underset{(3.24)}{6.37\ POPDEN} - \underset{(433.5)}{274.6\ \$\$} + \underset{(0.279)}{0.834\ REAL}$$

$$+ \underset{(0.003)}{0.025\ POP} + \underset{(2{,}913)}{16{,}980\ UPSMSA}$$

Where:

GLFLS = Number of adjusted fishing licenses sold in Great Lakes counties;

POPDEN = Population density measured in total population/square mile;

$$ = License cost;

REAL = Per capita income;

POP = Total county population;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more.

This model has an adjusted $r^2$ of .915 and a standard error of 4,187 licenses. The mean number of licenses sold for a sample of 187 cases was 17,158. This model has an error range of about 24% about the mean or 9% lower than the upstate fishing license sales model. The license cost variable, although left in the equation, is not statistically significant (different from zero).

The upstate SMSA dummy variable has a strong influence on fishing license sales, with almost 17,000 additional licenses being sold in those SMSA counties. As with the statewide fishing model, population and real income factors are strongly associated with fishing license sales.

# PART III:

## THE 1980 CROSS-SECTIONAL DATA BASE

The 1980 cross-sectional data base was established to examine the factors related to differences in the sale of licenses among counties in the most recent census year, 1980, when data were available for a wider variety of demographic variables. Because this data base covers one particular year when everyone pays the same fee for the same type of license, the license cost can not be a part of these models.

Demographic variables available for 1980 and examined in association with adjusted resident small game, big game, small game/big game combination, and fishing licenses sold were:

<u>Population</u>

Male, aged 14-17, 18-44, and 45-64

Male population per square mile

<u>Education</u>

Percent of males age 25 and over who graduated from high school

<u>Income</u>

Mean 1979 household income

<u>Employment/Unemployment</u>

Number employed, 16 years of age and over

Number unemployed, 16 years of age and over

Percent in agricultural employment

<u>Mobility/Stability</u>

Percent over 5 years of age living in the same county as in 1975

Family Stability

Percent aged 15-54 once married who in 1980 are divorced, widowed, or legally separated

Small Game License Sales

The best 1980 explanatory model developed for small game license sales for the cross-sectional data base (with standard errors in parentheses) was:

$$SGLS = \underset{(154)}{1390} + \underset{(0.003)}{0.073}\, P14\text{-}17\ \&\ 45\text{-}64 - \underset{(0.39)}{7.15}\, POPDEN$$

$$+ \underset{(0.001)}{0.005}\, AC - \underset{(27.68)}{113.02}\, PAg$$

Where:

SGLS = Number of adjusted small game licenses sold;

P14-17 & 45-64 = NY male population aged 14-17 and 45-64;

POPDEN = Population density measured in male population per square mile;

AC = Acres of cropland;

PAg = Percent employed in agriculture.

This model has an $r^2$ of .961, and an adjusted $r^2$ of .958. The five counties comprising New York City have been combined into one case, thus n=58. The standard deviation of licenses from the regression line is 520. The mean number of licenses sold is 2633 with least number sold in Hamilton County, 215, and the greatest number sold in Suffolk county, 11,295.

We examined the effect of total male population as well as that of age groups 14-17, 18-44, and 45-64 on small game license sales. The combination of the 2 age groups, 14-17 and 45-64, had the highest correlation (positive) with small game license sales.

The measure of population density in each county is negatively associated with small game license sales. Thus, in rural counties more licenses were sold per capita than in urban counties.

Acres of cropland, which is positively associated with license sales, is a surrogate measure for supply. This assumes that the greater the number of acres of cropland the more small game available for harvest. Thus this model indicates that counties with more cropland sell more licenses, other factors held constant.

The last statistically significant variable in the model is percent employed in agriculture. This variable was negatively correlated with license sales.

The population and population density variables are similar measures to those found in the longitudinal county-level data base. However, the 1980 data base provides more detailed information on population and the additional supply variable, acres of cropland. Other demographic variables, while sometimes significantly associated with small game license sales, did not add significantly to those in the model in explaining sales at the county level.

## Big Game License Sales

The best explanatory model developed for big game license sales (with standard errors in parentheses) was:

$$
\begin{aligned}
BGLS = &\underset{(14{,}641)}{-72{,}544} + \underset{(0.012)}{0.104}\ P18\text{-}44 - \underset{(0.25)}{1.09}\ POPDEN - \underset{(2{,}668)}{12{,}120}\ DOWN \\
&+ \underset{(107.9)}{559.0}\ ED + \underset{(122.7)}{487.0}\ NOMIGR + \underset{(0.009)}{0.020}\ AC + \underset{(0.56)}{1.12}\ BH\text{-}1
\end{aligned}
$$

Where:

BGLS = Number of adjusted big game licenses sold;

| | | |
|---|---|---|
| P18-44 | = | NY male population aged 18-44; |
| POPDEN | = | Population density measured in male population per square mile; |
| DOWN | = | Dummy variable indicating a downstate county (Westchester, Rockland, NYC, and Long Island); |
| ED | = | Percent of males, 25 years old and older, who have graduated from high school; |
| NOMIGR | = | Percent of people 5 years old and older who are living in the same county as in 1975; |
| AC | = | Acres of cropland; |
| BH-1 | = | Adult bucks harvested the previous year. |

This model has an $r^2$ of .765, and an adjusted $r^2$ of .735. Its standard error is 4036 or 39% of the mean number of licenses sold (10,348). The contrast in the least number sold, 1,427 in Hamilton county, and the greatest number sold, 48,878 in Erie county, illustrates the high variation in sales among counties and the difficulty in developing a highly accurate model for all counties in the state.

The male population aged 18-44 was the most highly correlated independent variable with big game license sales. Expansion to include those aged 45-64 did not improve the model.

The next two variables in the model relate to concentrations of populations. Both, population density and the downstate dummy variable, have negative coefficients. Thus rural, upstate counties contribute more to big game license sales than urban or downstate counties, other variables in the model held constant.

The next two variables in the model are demographic variables. Education had a positive association with license sales. This variable was highly correlated with income and may well reflect a higher standard of living and ability to purchase a license and hunting equipment. The second demographic variable is a measure of the stability of the population in a county. The higher the percent of people who are still living in the same county they lived in 5 years ago, the higher the license sales. This variable may imply some continuity in family and peer support structure which, based on other work done by the Human Dimensions Research Unit, is important to continuation in hunting (Purdy et al. 1985).

The last two variables in the model are measures of supply. Acres of cropland was positively associated with license sales, thus counties with cropland sell more licenses (all other factors held constant). Also, the number of adult bucks harvested the previous year was positively associated with big game license sales.

## Combination Small Game/Fishing License Sales

The best model that we could develop for the combination small game/ fishing licenses, given a very limited amount of fisheries data, is:

$$\text{SGFC} = \underset{(795)}{3{,}688} + \underset{(1{,}041)}{2{,}403}\ \text{GLC} - \underset{(1{,}719)}{6{,}823}\ \text{DOWN} + \underset{(0.006)}{0.031}\ \text{P14-44}$$

$$-\underset{(0.14)}{0.43}\ \text{POPDEN} - \underset{(138)}{503}\ \text{PAg} + \underset{(0.007)}{0.017}\ \text{AC}$$

Where:

SGFC = Number of adjusted small game/fishing combination licenses sold;

GLC = Dummy variable indicating counties bordering the Great Lakes;

DOWN = Dummy variable indicating a downstate county;

P14-44 = NY male population aged 14-44;

POPDEN = Population density measured in male population per square mile;

PAg = Percent employed in agriculture;

AC = Acres of cropland.

This model has an $r^2$ of .618, and an adjusted $r^2$ of .577. Its standard error is 2,554, over half the mean number of licenses sold of 4,349. Again, given the range of 669 sold in Richmond county versus 24,368 sold in Erie county, it is difficult to develop accurate models for all counties statewide.

The first variable in the model, a dummy variable for Great Lakes counties, predicts additional license sales of 2,403 for the 9 counties bordering the Great Lakes. As with big game license sales, the population density variable and the downstate dummy variable have negative coefficients for the small game/fishing license sales model. Thus, upstate, rural counties have greater license sales, holding population and other variables constant.

The best general population variable for predicting combination license sales was males aged 14-44. Percent employed in agriculture was negatively correlated with license sales. Acres of cropland is the last statistically significant variable in the model. It is a surrogate measure for small game supply and is positively correlated with license sales.

This model has a lower $r^2$ and higher proportional error than most previously discussed models, and there is also a lack of demographic variables in this model. This leads us to believe that all of the relevant variables which could explain these license sales have not been examined. Most likely these include supply variables relevant to fishing and small game.

Fishing License Sales

The best 1980 model that we could develop from the available data is:

$$\text{FLS} = \underset{(622)}{5,039} + \underset{(3,149)}{19,745}\ \text{EMOD} + \underset{(2,191)}{14,013}\ \text{UPSMSA} + \underset{(1,413)}{6075}\ \text{GLC} + \underset{(11.97)}{25.97}\ \text{INWATER}$$

Where:

FLS = Number of adjusted fishing licenses sold;

EMOD = Dummy variable representing Erie, Monroe, and Onondaga counties;

UPSMSA = Dummy variable representing the 6 upstate counties within each SMSA with a central city of 50,000 or more;

GLC = Dummy variable indicating counties bordering the Great Lakes;

INWATER = Square miles of inland water.

This model has an $r^2$ of .854, and an adjusted $r^2$ of .843. Its standard error is 3,678, or 41% of the mean number of licenses sold (8,948). The range in sales was from 571 in Richmond county to 51,416 in Erie county.

Multidimensional plots of fishing license sales with the independent variables consistently showed Erie, Monroe, and Onondaga counties to be clustered away from other counties; these 3 counties had higher sales and higher measures of other independent variables. Thus, this dummy variable was created for the fishing license sale model. This variable helps us to predict license sales but it does not help us to understand why these 3 counties in particular have such high license sales.

The second dummy variable in the model measures the positive effect of large upstate urban centers on fishing license sales. According to the model these counties sold an additional 14,000 licenses. Similarly, counties along the Great Lakes sold an additional 6,000 licenses due to their greater access to fishing waters.

The last variable in the model is also a measure of access and supply. The square miles of inland water is a rough approximation of the area available for fishing. It has a statistically significant, positive influence on fishing license sales. This variable points out the usefulness of this type of cross-sectional data base. The amount of inland water in New York State has not changed significantly in the last 20 years (i.e., no new reservoirs were built), so it is not an appropriate measure in the other two data sets, which measure license sales over time. But it is an important variable when trying to describe differences in fishing license sales between counties.

# IMPLICATIONS AND RECOMMENDATIONS

Work in the area of explaining and predicting license sales is very much an iterative process. One develops models from the data at hand, analyzes them, and tries to improve them. We now have 2 years experience with the statewide models, and have produced models from 3 of the 4 license types that are quite accurate. More work needs to be done with the projection of fishing licenses. We have only 1 year of experience with the 2 county-level data sets. Quite a bit of additional work is needed with these models. This will be detailed after a summary of findings and implications.

This work began with the general objective of trying to predict the magnitude and direction of license sales and the effect of such sales on revenues. At the statewide planning level, the statewide longitudinal data base is the best data base for projections. Models from that data base have consistently shown (1) expected annual growth in license sales and (2) that increases in license fees will increase revenues. Despite fluctuations in license sales that resulted in some declines in 1984-85, actual sales were within the error range of models developed through 1983-84. Inserting 1984-85 data resulted in minor adjustments of coefficients but yielded the same implications of increased sales in future years.

Of the statewide models, only the model for fishing license sales seems questionable. With the input of 1984-85 data, no population variable is statistically significant, nor is the license fee. The overall model is highly significant, and the prediction error rate is only about 5%, but a model with no population and no resource variables is suspect. The fact that the license fee is not significant likely means simply that an additional fee of moderate amount would not affect sales.

Although the statewide longitudinal data base is theoretically the best data base for statewide sales and revenue projections, it has limitations. Relatively few demographic variables have annual data available for 20 to 25 years. Only for big game do we appear to have adequate resource data. The statewide longitudinal data base has only about 20 data points -- enough statistically, but perhaps not enough to generate confidence that the effects measured are stable. And although residual analysis shows no notable differences in projection errors for early versus recent years, there is cause to wonder whether sales, demographic, and resource relationships of the early 1960's still hold today.

For these reasons, 2 additional data bases were formed and examined, the countywide longitudinal data base from 1963 to 1983, and the county cross-sectional data base for 1980. Demographic variables for these models were not always identical to those for the statewide model, but the county-level models generally confirmed that independent variables found to be statistically significant statewide were also significant in explaining differences in sales across counties. Some additional insights came from these data bases. Population density has a negative effect on most license sales, and more hunting licenses per capita are sold in counties where people move less frequently.

This progress report was prepared at this time because it has been a year since the previous report, and because of DEC requests for additional progress. Additional work is needed in particular to continue analyzing the potential usefulness of regional models. We recommend the following for the coming year:

# LITERATURE CITED

Brown, T.L. 1985. Development of Explanatory and Predictive Models for Hunting and Fishing License Sales and Revenue Trends in New York. Outdoor Recreation Research Unit Publ. 85-1, Dept. of Nat. Resour., N.Y.S. Col. of Agr. and Life Sci., Cornell Univ., Ithaca. 35 pp.

Purdy, K.G., D.J. Decker and T.L. Brown. 1985. New York's Hunter Training Course Participants: The Influence of Social-Psychological Influences on Participation in Hunting from 1978-1984. Human Dimensions Research Unit Publ. 85-7, Dept. of Nat. Resour., N.Y.S. Col. of Agr. and Life Sci., Cornell Univ., Ithaca. 127 pp.