

## A Computer Program for Age–Length Keys Incorporating Age Assignment to Individual Fish

DANIEL A. ISERMANN<sup>\*1</sup> AND CAREY T. KNIGHT

*Ohio Department of Natural Resources, Division of Wildlife, Fairport Fisheries Research Station,  
1190 High Street, Fairport Harbor, Ohio 44077, USA*

**Abstract.**—We provide a computer program (AGE-KEY) that quickly implements an age–length key approach for estimating fish population age structure and mean lengths at age. The program’s approach to age assignment differs from previous approaches in that **ages are explicitly assigned to individual unaged fish**. Because individual age assignment is necessary to describe the variation around estimates of mean catch per effort (CPE) at age and to compare these estimates through standard statistical procedures, the number of fish of a specified age captured in each unit of effort must be known or estimated. Individual age assignment can be confounded by instances of fractionality, where ages cannot be evenly assigned to fish within individual length intervals. Program AGEKEY resolves the problem of fractionality by applying an **integer-based approach to age assignment**. When we tested the validity of the program with two sample data sets, AGEKEY yielded estimates of age frequency and mean lengths at age that were very similar to estimates calculated through spreadsheet approaches. Program AGEKEY can quickly derive estimates of age frequency and mean lengths at age using an age–length key approach that will provide users with the data necessary to describe the precision associated with estimates of CPE at age and allow for statistical comparisons of CPE at age across surveys.

Because time constraints make it difficult to directly age all fish in a sample, biologists commonly attempt to collect uniform subsamples within specified length strata (i.e., 10 fish per 10- or 25-mm length intervals) to estimate the age frequencies of fish populations. When using a stratified approach, correct estimation of age frequency, growth, and annual mortality requires that ages be extrapolated to all fish within a sample, which is frequently accomplished with an age–length key (Hoenig and Heisey 1987; DeVries and Frie 1996; Bettoli and Miranda 2001). Although the use of age–length keys is a relatively simple concept,

construction and application of numerous age–length keys can be time-consuming. Some computer programs such as FishCalc89-DisBCal89 (developed by the Missouri Department of Conservation and available through the American Fisheries Society Computer User Section; available at [www.fisheries.org/cus/](http://www.fisheries.org/cus/)) and WINFIN (developed by the Nebraska Game and Parks Commission; available at [www.ngpc.state.ne.us](http://www.ngpc.state.ne.us)) offer the ability to implement age–length keys, but it can be difficult to accommodate varying length intervals used for subsampling. We also know that many biologists use various spreadsheet applications to construct and implement age–length keys; however, spreadsheet approaches often require a customized configuration to handle varying length distributions encountered across species.

Age frequency and mean lengths at age can be estimated with an age–length key approach without explicitly assigning ages to individual unaged fish within a sample. Furthermore, after the age frequency of a stratified subsample has been extrapolated to the entire sample, estimates of mean catch per effort (CPE) at each age can be calculated by dividing the total number of fish captured at a specified age by the total units of effort (i.e., number of gill-net sets). However, **assigning ages to individual fish is a necessary step in describing variation associated with estimates of CPE at age**. This is because the number of fish of a specified age in each individual unit of effort (i.e., individual gill-net sets) must be known to calculate standard measures of data dispersion (i.e., variance, standard error, coefficient of variation) and is also needed to make comparisons of CPE at age across samples using standard statistical procedures (i.e., *t*-tests, analysis of variance [ANOVA]) that operate by partitioning variance (i.e., sums of squares).

We present a computer program that quickly constructs and implements age–length keys. The program can easily accommodate different length intervals used for subsampling and does not need to be reconfigured to reflect differences in length

\* Corresponding author: [dan.isermann@dnr.state.mn.us](mailto:dan.isermann@dnr.state.mn.us)

<sup>1</sup> Present address: Minnesota Department of Natural Resources, Section of Fisheries, 1601 Minnesota Drive, Brainerd, Minnesota 56401, USA.

Received June 29, 2004; accepted March 7, 2005

Published online August 3, 2005

distributions observed across species. The program's approach to age assignment differs from previous approaches in that **ages are explicitly assigned to individual unaged fish in a sample**, providing the user with the data necessary to calculate precision associated with estimates of CPE at age and to compare these estimates across surveys. Lastly, because the program provides an **age for each fish within the entire sample**, **mean lengths at age** and associated **variance** can be **estimated** through standard equations.

### Methods

*Age-length key assumptions.*—Standard use of the classic age-length key approach first presented by Fridriksson (1934; cited in Ketchen 1949) adheres to the primary assumption that the length-based age distributions observed in the subsample of aged fish are representative of the distributions in the unaged sample to which ages are to be extrapolated (Westrheim and Ricker 1978). Numerous factors can result in a violation of these conditions; these factors include spatial and temporal differences in age distribution or growth among samples of fish and differences in selectivity among sampling gears. To adhere to the assumption of constant age composition, we suggest that **age-length keys be constructed with fish that have been collected simultaneously following the same sampling technique** along with proper consideration for spatial differences in growth and age composition. In instances where the assumption that length-based age distributions are constant across samples may be violated, we direct readers to the work of Hoenig and Heisey (1987), who present a maximum-likelihood approach to age-length keys.

*Program description.*—We constructed the age-length key program (AGEKEY) in the SAS programming language (SAS Institute 2000) because SAS software is commonly used to analyze fishery data. Program AGEKEY (Appendix) consists of two separate program loops. In the first loop (LOOP 1), the **overall age frequency** for an **individual length interval** is estimated by extrapolating the interval-specific age distribution observed in the aged subsample to the total number of unaged fish collected in the specified interval. For example, if 10 of 100 fish within a length interval of 225–249 mm were aged, and 50% of the aged fish ( $N = 5$ ) were age 3 and 50% ( $N = 5$ ) were age 4, LOOP 1 would randomly assign ages to the unaged fish within a length interval based on the estimated age distribution for that interval. In other

words, LOOP 1 would randomly assign ages to unaged fish in the 225- to 249-mm length interval so that 45 of the unaged fish would be assigned an age of 3 and 45 fish would be assigned an age of 4.

For length intervals where ages can be evenly applied to unaged fish (as in the previous example), the distribution of ages assigned to unaged fish will exactly match the age distribution observed from the subsample of aged fish. However, the observed distribution of ages within a length interval often **cannot be evenly applied** to unaged fish within the length interval, a circumstance we refer to as **fractionality**. For example, if 43 unaged fish were present in a 25-mm length interval of 225–249 mm, and the age distribution observed from the aged subsample is divided evenly between fish ages 3 and 4, then 21.5 of the unaged fish should be assigned an age of 3 and the remaining 21.5 of the fish should be assigned an age of 4. Because it is not reasonable that 21.5 fish of a specified age were caught in the sample, fractionality can result in minor abnormalities when reporting raw age frequencies. In some cases, rounding may be applied (DeVries and Frie 1996) so that age frequencies can be reported as whole numbers.

To cope with the problem of fractionality, LOOP 1 **assigns ages on an integer basis only**. Therefore, in the previous example 21 fish would be randomly assigned an age of 3 and 21 fish would be assigned an age of 4, leaving 1 fish with no age assignment. Fish not assigned an age in LOOP 1 are treated as a new sample of unaged fish in LOOP 2. In LOOP 2, **ages are assigned** to the remaining unaged fish by **randomly sampling from an array that reflects the distribution of observed ages for a specific length-group**, similar to the assignment process used in LOOP 1. Data for aged fish and fish that were assigned ages are then merged into an output data file (AGEFINAL).

*Data compatibility and array sizes.*—The **input data set must include both aged and unaged fish**. As presented, the input data set that contains all aged and unaged fish is called ALLFISH. To ensure compatibility with the program as it is presented, input data must be numerical, in columnar format, and must include at least three columns that denote (1) fish length in mm (variable name = TL); (2) sex of the fish (variable name = SEX); and (3) age of the fish (variable name = AGE). A subset of hypothetical sample data is presented with the program (Appendix). Additional data columns can be included (i.e., weight, sample loca-

tion), but must be denoted in the input statement of the program. The order of data columns must match the order of variable names listed in the input statement appearing in the beginning lines of the program. It should be noted that, by including a data column that denotes the sample that each fish was captured in (as in the yellow perch *Perca flavescens* data included in the Appendix), the output data set AGEFINAL can be used to calculate catch-at-age data for each individual unit of sampling effort. This information, combined with any samples that resulted in zero catches, can subsequently be used to calculate CPE at age and associated measures of variability and to make comparisons of CPE at age across surveys.

The length intervals (variable name = LI) used when collecting stratified samples of fish for age determination can be defined by the user in a separate column (denoted LI) during data entry or can be defined directly in the first few lines of the program. For example, the program presented in the Appendix is configured for 25-mm length intervals; 10-mm length intervals can be accommodated by replacing 25 with 10 on the designated line of the program (Appendix). Unaged fish within a sample must be denoted by entering a “.” in the proper column. If sex-specific age assignment is not desired or sexes are not recorded, the variable name SEX can be deleted in all instances where it appears in the program or all fish can be designated with a similar code for sex in the input data set. In cases where sex-specific age assignment is desired, fish that are not sexed should be denoted with a “.” in the appropriate column and these fish will not be assigned an age.

Array sizes in the program are currently set to handle situations where (1) the maximum number of unaged fish in any specified length interval does not exceed 1,000, (2) the maximum number of ages observed within any length interval does not exceed 20, and (3) the maximum number of fish in any length interval that are not assigned ages in LOOP 1 does not exceed 50. We expect that the above conditions will accommodate most age-length key applications and that relatively few fish without assigned ages will typically remain after LOOP 1 is completed (<10 fish in all of our trial runs of the program). Users can adjust the program for situations where the above conditions are not met; the portions of the code that would need to be adjusted are denoted in the program.

**Program validation.**—We tested the validity of program AGEKEY using two examples of varying complexity. We first applied our program to data

used by DeVries and Frie (1996) to describe age structure of white crappies *Pomoxis annularis*. White crappies had been collected by means of a stratified sampling design, where a maximum of 10 fish were randomly sampled from within each 10-mm length interval for age determination. Age structure for the entire sample of white crappies was estimated with an age-length key. Problems with fractionality were resolved by rounding as previously described. The white crappie example offered a simple case where sex-specific age structure was not of interest and variation in ages within individual length intervals was relatively low (DeVries and Frie 1996). Because sex was not identified in this data set, all white crappies were given a similar code for sex (SEX = 1) during data entry. Age frequencies generated through a spreadsheet approach were compared with results generated through program AGEKEY. Age frequencies resulting from the spreadsheet approach were rounded to match those reported by DeVries and Frie (1996).

Age- and length-frequency data recorded from a sample of yellow perch ( $N = 922$ ) collected in 68 bottom trawls from Lake Erie during August 2003 were used as an additional test of program AGEKEY. A subset of this data set is included within the program as it is presented in the Appendix. This example was more complex in that sex-specific age assignment was necessary and age variation within specific length intervals was relatively high. All yellow perch in the sample were measured to the nearest millimeter (total length) and sex was identified for each fish through an inspection of gonads. In this example, an additional data column was included to denote the individual trawling run in which each yellow perch was captured (SAMPLEID; Appendix) so that catch-at-age data for each trawl were available.

The sampling objective was to remove otoliths from a maximum of 50 yellow perch of each sex within each 25-mm length interval for age determination. Only yellow perch 125 mm and larger were aged. All female yellow perch 300 mm or larger and all male perch 250 mm or larger were aged. Before completing program AGEKEY, sex-specific age-length keys (Table 1) were constructed and ages were extrapolated to the unaged samples of fish with Microsoft Excel (Microsoft 1999) spreadsheets. Age frequencies generated through the spreadsheet approach were compared with results generated through program AGEKEY. Mean length at age  $i$  ( $\bar{L}_i$ ) and the associated variance ( $S_i^2$ ) were calculated from spreadsheet results with the

TABLE 1.—Estimated mean lengths at age (mm) and SEs (in parentheses) for female (F) and male (M) yellow perch collected in bottom trawls from the central basin of Lake Erie in August 2003. Mean lengths and SEs were estimated from age-length key data with approaches presented by DeVries and Frie (1996; denoted as MPA) and Bettoli and Miranda (2001; denoted as MLA), and program AGEKEY.

Sex	Method	Age						
		1	2	3	4	5	6	7
F	MPA	150 (3.5)	180 (1.0)	224 (4.5)	251 (2.7)	273 (5.7)	275 (24.0)	301 (14.0)
	MLA	150 (5.0)	179 (1.0)	224 (4.0)	252 (2.6)	268 (6.4)	279 (23.0)	304 (12.0)
	AGEKEY	145 (4.0)	180 (1.0)	219 (4.8)	251 (2.6)	269 (6.5)	276 (21.0)	306 (14.0)
M	MPA	138 (0.0)	170 (1.2)	214 (2.5)	228 (1.7)	229 (3.2)	237 (3.5)	241 (5.6)
	MLA	137 (2.5)	169 (1.1)	213 (1.8)	227 (1.7)	228 (3.0)	238 (4.2)	240 (5.2)
	AGEKEY	142 (2.5)	169 (1.1)	214 (2.6)	225 (1.5)	231 (3.9)	240 (5.2)	240 (6.9)

following equations (Bettoli and Miranda 2001; see equations 3 and 4a; denoted as MLA):

$$\bar{L}_i = \frac{\sum N_{ij} \bar{l}_{ij}}{N_i} \quad \text{and} \quad (1)$$

$$S_i^2 = \frac{\sum N_{ij} \left[ \sum \frac{(l_{ijk} - \bar{L}_i)^2}{n_{ij}} \right]}{N_i - 1}; \quad (2)$$

$l_{ijk}$  = the length of the  $k$ th age  $i$  fish in the  $j$ th length-group within the subsample of aged fish;

$N_{ij}$  =  $N_j(n_{ij}/n_j)$ ;

$N_j$  = the number of fish in the  $j$ th length-group;

$n_j$  = the total number of fish in the aged subsample from the  $j$ th length-group;

$n_{ij}$  = the number of age  $i$  fish in the aged subsample from the  $j$ th length-group;

$N_i$  =  $\sum N_{ij}$  over all  $j$  length-groups; and

$\bar{l}_{ij}$  = the mean length of age  $i$  fish in the  $j$ th length-group.

We also calculated  $\bar{L}_i$  using the approach outlined by DeVries and Frie (1996, denoted as MPA) by replacing  $\bar{l}_{ij}$  in equation (1) with the midpoint of each  $j$ th length interval ( $m_j$ ). Variance for MPA estimates of mean length at age were estimated by

$$S_i^2 = \frac{\sum N_{ij}(m_j - \bar{L}_i)^2}{N_i - 1}, \quad (3)$$

an equation similar to equation (4b) presented by Bettoli and Miranda (2001). Standard errors were calculated as  $SE = [S_i^2/N_i]^{0.5}$  when both MLA and MPA were used. Mean lengths at age and standard errors resulting from MLA and MPA were compared against estimates generated by program AGEKEY.

## Results and Discussion

The age frequency reported by DeVries and Frie (1996) for white crappies differed only slightly (differences of only a single fish at each age) from the distribution generated by program AGEKEY. The slight differences in individual age frequencies resulted from the random age assignment in LOOP 2 of program AGEKEY (rather than rounding) for estimating age frequencies when fractionality occurred.

Fractionality occurred in nearly all length intervals for both male and female yellow perch collected from Lake Erie (Figure 1). For most ages, age frequencies generated from program AGEKEY differed only slightly (a difference of >3 fish in only one instance) from age frequencies generated with the spreadsheet approach (Figure 1). Again, the slight variation in age frequencies between the two methods was a result of program AGEKEY's approach to resolving the problem of fractionality.

Differences among yellow perch mean lengths at age estimated by MLA, MPA, and program AGEKEY were 5 mm or less (Table 1), and differences in the magnitude of SE were 2 or less across all ages. Obviously, we cannot say which approach most accurately reflects the estimates of mean lengths at age and SE that would have been generated had the ages of all fish in the sample been known, but each of the approaches has obvious drawbacks. If the true mean length of age  $i$  fish in a specific length interval is not equivalent to the midpoint of the interval, then subsequent estimates of mean lengths at age generated with MPA will be biased to some degree. Furthermore, problems arise when estimating variance with MPA when all fish of a given age are collected within one length interval, as  $m_j - \bar{L}_i$  in equation (3) equals zero and results in an estimate of no variance (i.e., age-1 male

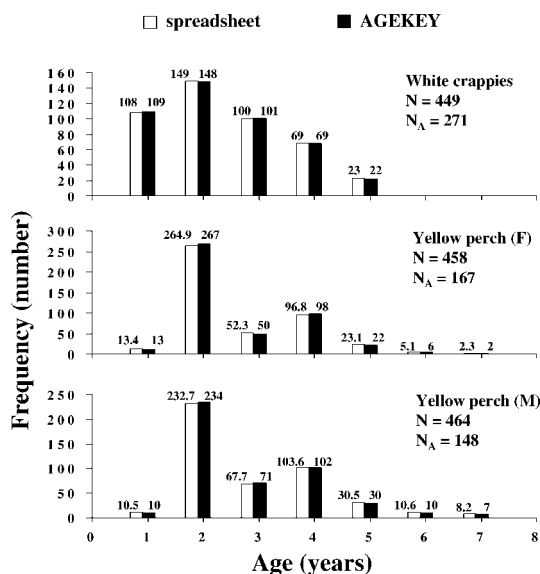


FIGURE 1.—Age frequencies for white crappies (DeVries and Frie 1996) and Lake Erie yellow perch (females [F] and males [M]) generated from a spreadsheet application of age–length keys and the program AGEKEY. The total number of fish sampled ( $N$ ) and size of the subsample that were aged ( $N_A$ ) are reported. White crappie age frequencies generated through a spreadsheet approach were rounded in order to match those reported by DeVries and Frie (1996).

yellow perch; Table 1). A foreseeable problem with the MLA approach relates to the fact that, within the aged subsample the number of age  $i$  fish in a given interval is often very low (sometimes only a single fish), possibly providing a relatively poor estimate of the actual mean length of age  $i$  fish in the specified interval ( $\bar{L}_{ij}$ ). The assignment of ages to individual fish in program AGEKEY may result in increased sample sizes for estimating the mean length of age  $i$  fish in a length interval, but some of the age assignments may be wrong and result in biased estimates. Regardless of the approach, the size of the length intervals used for subsampling and the number of fish that are aged within each interval will greatly affect the accuracy and precision of mean length-at-age estimates. **Increasing the number of fish that are aged within each length interval or reducing interval size will provide improved mean length-at-age estimates.** In cases where the length interval used for subsampling exceeds 25 mm, we suggest that MLA or MPA would be more reasonable approaches for estimating mean lengths at age than program AGEKEY because the probability of incorrectly assigning ages to individual fish would be higher.

Additional tests of program AGEKEY using data from several species have yielded similar favorable results. Troubleshooting the program has been relatively easy, as errors result in unaged fish in the AGEFINAL data set. In all of our trial runs, errors have stemmed from data problems rather than from problems with the program code. The most common problem we have encountered has been when unaged fish occur in length intervals that contain no aged fish, which represents an error that occurred at the time aging structures were collected.

### Acknowledgments

Program AGEKEY could not have been completed without the assistance of personnel at SAS Technical Support, who requested to remain anonymous. Comments from M. Hansen, S. Banks, P. Hanchin, D. Kirby, C. Vandergoot, and several anonymous reviewers improved the manuscript. We also thank P. Hanchin (Michigan Department of Natural Resources), C. Vandergoot (Ohio Division of Wildlife), and G. Scholten (Tennessee Tech University) for supplying additional data sets used to test the program.

### References

- Bettoli, P. W., and L. E. Miranda. 2001. A cautionary note about estimating mean length at age with subsampled data. *North American Journal of Fisheries Management* 21:425–428.
- DeVries, D. R., and R. V. Frie. 1996. Determination of age and growth. Pages 483–512 in B. R. Murphy and D. W. Willis, editors. *Fisheries techniques*, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Fridriksson, A. 1934. On the calculation of age distribution within a stock of cod by means of relatively few age determinations as a key to measurements on a large scale. *Rapports et Procès-Verbaux des Réunions, Conseil International pour l'Exploration de la Mer* 86(6).
- Hoening, J. M., and D. M. Heisey. 1987. Use of a log-linear model with the EM algorithm to correct estimates of stock composition and to convert length to age. *Transactions of the American Fisheries Society* 116:232–243.
- Ketchen, K. S. 1949. Stratified subsampling for determining age distributions. *Transactions of the American Fisheries Society* 79:205–212.
- Microsoft Corporation. 1999. *Microsoft Excel*. Microsoft Corporation, Redmond, Washington.
- SAS Institute. 2000. *The SAS system*, version 8.1, SAS Institute, Cary, North Carolina.
- Westrheim, S. J., and W. E. Ricker. 1978. Bias in using an age–length key to estimate age-frequency distributions. *Journal of the Fisheries Research Board of Canada* 35:184–189.



**Appendix: Program AGEKEY**

Program AGEKEY is presented in the SAS programming language (SAS Institute 2000). The program is currently set up to handle 25-mm length intervals; different length intervals used for subsampling must be specified by the user where denoted in the program. A subset of yellow perch data is included as an example of proper data formatting. Electronic versions of the program can be obtained by contacting the corresponding author.

```
DATA ALLFISH (INDEX = (SEXLI = (SEX
  LI))) AGEMISS;
/* NEXT STATEMENT DEFINES THE VARIABLES
  APPEARING IN DATA COLUMNS */
INPUT SAMPLEID MM SEX AGE;
/* NEXT DEFINE DESIRED LENGTH INTERVALS
  (LI) IN MM. CURRENTLY SET UP FOR 25-
  MM INTERVALS. IF 25-MM INTERVALS ARE
  NOT DESIRED, USER MUST ADJUST BY RE-
  PLACING 25'S WITH THE DESIRED INTER-
  VAL WIDTH*/
LI = (INT(MM/25))*25;
/* CREATE DATA SET AGEMISS THAT INCLUDES
  ONLY THE UNAGED FISH*/
IF AGE EQ. AND SEX NE. THEN OUTPUT AGEM-
  ISS;
/* CREATE DATA SET ALLFISH THAT INCLUDES
  ALL FISH (AGED AND UNAGED) */
OUTPUT ALLFISH;
/* BELOW THE CARDS STATEMENT IS A SUBSET
  OF THE DATA USED IN THE YELLOW PERCH
  EXAMPLE THAT INCLUDES ONLY THE MALE
  PERCH BETWEEN 126 AND 149 MM TL*/
CARDS;
1 127 1.
1 129 1.
1 130 1.
2 132 1.
2 132 1.
3 134 1.
4 134 1.
5 137 1.
6 142 1.
11 143 1.
19 143 1.
25 144 1.
32 145 1.
44 146 1.
44 147 1.
48 147 1.
49 147 1.
```

```
50 148 1.
67 149 1.
1 126 1 1
3 131 1 1
5 141 1 1
9 141 1 1
12 148 1 1
31 126 1 2
31 136 1 2
37 136 1 2
38 140 1 2
39 140 1 2
44 141 1 2
47 144 1 2
53 145 1 2
57 148 1 2
62 148 1 2
66 149 1 2
68 149 1 2
;
/* SORT THE DATA SET BY SEX AND LENGTH
  INTERVAL */
PROC SORT DATA = AGEMISS;
BY SEX LI;
RUN;
/* TALLY THE NUMBER OF UNAGED FISH IN
  EACH LENGTH INTERVAL
  AND MERGE WITH THE AGEMISS DATA SET, SO
  THAT FOR EACH UNAGED FISH THE TOTAL
  NUMBER OF UNAGED FISH IN THAT LENGTH
  INTERVAL IS KNOWN */
PROC FREQ DATA = AGEMISS NOPRINT;
TABLES SEX*LI/OUT = OUT (DROP = PER-
  CENT);
RUN;
DATA AGEMISS;
MERGE AGEMISS OUT;
BY SEX LI;
RUN;
/* CREATE A DATA SET NAMED AGEDFISH,
  WHICH CONTAINS ONLY THE AGED FISH */

DATA AGEDFISH;
SET ALLFISH;
IF SEX>0;
IF AGE = . THEN DELETE;
RUN;
PROC SORT DATA = AGEDFISH;
BY SEX LI AGE;
RUN;
/* TALLY THE NUMBER OF AGED FISH AT EACH
```

```

    AGE IN EACH LENGTH INTERVAL FOR EACH
    SEX */
PROC MEANS DATA = AGEDFISH N NOPRINT;
BY SEX LI AGE;
VAR MM;
OUTPUT OUT = NEW(DROP = _TYPE_ _FREQ_ IN-
    DEX = (SEXLI = (SEX LI))) N = NUM-
    BER;
RUN;
/* TALLY THE TOTAL NUMBER OF AGED FISH IN
    EACH LENGTH INTERVAL FOR EACH SEX */

PROC MEANS DATA = NEW SUM NOPRINT;
BY SEX LI;
VAR NUMBER;
OUTPUT OUT = NEW2(DROP = _TYPE_ _FREQ_
    INDEX = (SEXLI = (SEX LI))) SUM =
    TOTAL
RUN;
/* START OF LOOP 1--ASSIGNS AGE VALUES TO
    UNAGED FISH ON AN INTEGER BASIS.
    FIRST STEP IS TO DEFINE THE MAX NUM-
    BER OF AGES AND UNAGED FISH IN ANY
    LENGTH INTERVAL FOR CASES > 1000.
    USER MUST ADJUST MAX SIZE */
%LET MAX = 1000;
DATA AGEMISS;
/* THE ARRAYS THAT WILL BE FILLED ARE
    BASED ON THE NUMBER OF UNAGED FISH IN
    A LENGTH INTERVAL AND THE NUMBER OF
    AGES OBSERVED IN THE INTERVAL*/
ARRAY VALUES (&MAX) (&MAX*.);
ARRAY COUNTS (&MAX) (&MAX*.);
/* ADDED THESE TO RANDOMIZE THE VALUES
    IN THE ARRAYS. */
ARRAY RANDOM (&MAX) (&MAX*.);
RETAIN RANDOMIZED;
/* READ AN OBSERVATION FROM AGEMISS. */
MODIFY AGEMISS;
/* NEEDS AGE FILLED IN. */
IF (SEX NE. ) AND (AGE EQ. ) THEN DO;
/* */
/* CREATE LAGGED VARIABLES TO DETECT DU-
    Plicate SEX/LI VALUES. */
/* */
LAG_SEX = LAG(SEX);
LAG_LI = LAG(LI);
/* */
/* CLEAR OUT THE ARRAYS, BUT ONLY FOR
    NONDUPLICATE SEX/LI VALUES. */
/* */
IF NOT (LAG_LI EQ LI AND LAG_SEX EQ SEX)
    THEN DO;
    BYCOUNT = 0;
    READINDEX = 1;
    RANDOMIZED = 0;
    DO I = 1 TO DIM(VALUES);
        VALUES(I) = . ;
        COUNTS(I) = . ;
    END;
    END;
    BYCOUNT + 1;
    /* */
    /* FETCH MATCHING OBSERVATION FROM NEW2
        TO GET TOTAL. */
    SET NEW2 KEY = SEXLI;
    FILLINDEX = 1;
    /* FETCH ALL MATCHING OBSERVATIONS FROM
        NEW TO GET SAMPLE AGE */
    /* VALUES AND COUNTS. */
    DO UNTIL (_IORC.);
        SET NEW KEY = SEXLI;
        /* USE AGE VALUES FROM MATCHING OBSER-
            VATIONS TO FILL VALUES ARRAY. */
        /* NOTE ADDITION OF _IORC_ TEST. IF SEX/
            LI IS A DUPLICATE, THE */
        /* ARRAY ALREADY CONTAINS THE CORRECT
            VALUES TO SAMPLE FROM. */
        IF NOT (_IORC_) THEN DO;
            VALUES(FILLINDEX) = AGE;
            COUNTS(FILLINDEX) = NUMBER;
            FILLINDEX + 1;
        END;
        /* ONCE THE VALUES AND COUNTS ARRAYS ARE
            FULL, RANDOMIZE THEM */
        /* BUT ONLY ONCE, JUST AFTER THEY'VE
            BEEN POPULATED. */
        IF _IORC_ AND NOT RANDOMIZED THEN DO;
            RANDOMIZED = 1;
            DO I = 1 TO N(OF VALUES(*));
                RANDOM(I) = RANUNI(54321);
            END;
            DO UNTIL (SORTED);
                SORTED = 1;
                DO I = 1 TO N(OF VALUES(*))-1;
                    IF RANDOM(I)>RANDOM(I + 1) THEN DO;
                        TEMP = RANDOM(I + 1);
                        TEMP2 = VALUES(I + 1);
                        TEMP3 = COUNTS(I + 1);
                        RANDOM(I + 1) = RANDOM(I);
                        VALUES(I + 1) = VALUES(I);
                        COUNTS(I + 1) = COUNTS(I);
                        RANDOM(I) = TEMP;
                        VALUES(I) = TEMP2;
                        COUNTS(I) = TEMP3;
                    END;
                END;
                SORTED = 0;
            END;
            END;

```

```

END;
END;
END;
IF BYCOUNT GT INT((COUNTS(READINDEX)/
    TOTAL)*COUNT) THEN DO;
    READINDEX + 1;
    BYCOUNT = 1;
END;
AGE = VALUES(READINDEX);
/* */
/* REPLACE THE OBSERVATION WITH THE NON-
    MISSING AGE VALUE. */
/* */
REPLACE AGEMISS;
/* */
/* CLEAN UP. */
/* */
_IORC_ = 0;
_ERROR_ = 0;
END;
RUN;
/* LOOP 2--ASSIGNS AGES TO FISH THAT WERE
    NOT ASSIGNED AN AGE IN LOOP 1 DUE TO
    FRACTIONALITY. ASSIGNS AGES IN A
    MANNER SIMILAR TO THAT USED IN LOOP 1
    */
DATA AGEDFISH; SET ALLFISH;
IF SEX>0;
IF AGE = . THEN DELETE;
RUN;
PROC SORT DATA = AGEDFISH; BY SEX LI
    AGE; RUN;
PROC MEANS DATA = AGEDFISH N NOPRINT;
BY SEX LI AGE; VAR MM;
OUTPUT OUT = NEW(DROP = _TYPE_ _FREQ_ IN-
    DEX = (SEXLI = (SEX LI))) N = NUM-
    BER;
PROC MEANS DATA = NEW SUM NOPRINT ;
BY SEX LI; VAR NUMBER;
OUTPUT OUT = NEW2(DROP = _TYPE_ _FREQ_
    INDEX = (SEXLI = (SEX LI))) SUM =
    TOTAL;
DATA AGEMISS;
/* SET ARRAY RANDOM TO HANDLE CASES
    WHERE THE NUMBER OF UNAGED FISH RE-
    MAINING AFTER LOOP 1 IS 50 FISH OR
    LESS, USER MUST ADJUST BOTH VALUES
    FOR CASES > 50 */
ARRAY RANDOM(50)(50*.);
MODIFY AGEMISS;
IF (SEX NE. ) AND (AGE EQ. ) THEN DO;
    LAG_SEX = LAG(SEX); LAG_LI = LAG(LI);
    IF NOT (LAG_LI EQ LI AND LAG_SEX EQ SEX)
        THEN
        DO I = 1 TO DIM(RANDOM); RANDOM(I) = . ;
            END;
    SET NEW2 KEY = SEXLI; INDEX = 1; DO UNTIL
        (_IORC_);
    SET NEW KEY = SEXLI;
    IF NOT (_IORC_) THEN DO I = 1 TO NUMBER;
        RANDOM(INDEX) = AGE; INDEX = INDEX + 1;
    END; END;
    AGE =
        RAN-
        DOM(CEIL(RANUNI(54321)*TOTAL));
    REPLACE AGEMISS; _IORC_ = 0; _ERROR_ = 0;
    END; RUN;
/* NOW ALL FISH HAVE AN AGE. MERGE DATA
    SETS AGEDFISH AND AGEMISS INTO DATA
    SET AGEFINAL THAT INCLUDES ALL OF
    THE FISH WITH THEIR AGES */
DATA AGEFINAL; SET AGEDFISH AGEMISS;
DROP COUNT;
/* CODE TO OUTPUT AGE FREQUENCIES BY SEX
    */
PROC SORT; BY SEX;
PROC FREQ; TABLES LI*AGE; BY SEX; RUN;
/* CODE TO OUTPUT MEAN LENGTHS AT AGE BY
    SEX */
PROC SORT; BY SEX AGE;
PROC MEANS N MEAN STDERR; VAR MM; BY SEX
    AGE; RUN;
/* CODE TO OUTPUT NUMBER OF FISH OF SPE-
    CIFIC SEX AND AGE CAPTURED IN EACH
    UNIT OF SAMPLING EFFORT. NOTE: DOES
    NOT INCLUDE SAMPLES WHERE TOTAL
    CATCH = 0. */
PROC SORT; BY SAMPLEID SEX AGE;
PROC FREQ; TABLES SAMPLEID*AGE; BY SEX;
    RUN;

```