

QA/QC procedures used for TIP Online data

Dave Gloeckner
NMFS SEFSC
01/18/06

The TIP online application uses checks against code tables to capture out of range values for states, counties, gears, areas, weight units, grades, length units and length types.

The data was checked for invalid state, county, gear, weight units, grade, length units and length types using SPSS or SAS to further identify erroneous codes.

Effort data often contains more than one gear or area code for a given interview. The join with the effort table often results in a cartesian product producing duplicate records. The previous version of TIP required a duplicate entry for gear when a sample could not be tied to market category, but only the species. In this case an uncoded gear appears in the same interview with a valid gear. These duplications were identified and removed. Other duplications were assigned a single gear based on the gear with the greater soak time. If the soak times were equal or not recorded the gear was assigned based on the predominant gear for that cell, based on the total number of samples. After these duplications are removed, the effort data is joined with the sample data.

Lengths were all converted to total length in cm, while weights were converted to whole weight in kilograms. This was accomplished using conversion tables available from the SEFSC database and using published conversions.

Lengths that were out of range, > 145 cm TL, were then removed from the data to attempt to eliminate further miscoded data resulting from erroneous but valid length units, or possibly species misidentification.

Weights or weight units were only supplied for 8,649 record of the 118,125 records used in the final data set, with highly variable values. The variability was most likely the result of differing equipment or methods used for obtaining weights. For these reasons all weights were standardized using the TL to Whole Weight conversion in Schirripa and Goodyear (1994).