

2021 American Samoa Small Boat Fishery Cost-Earnings Survey

Title: Data cleaning documentation

Author: Crystal Dombrow

Date: June - September 2022

Description: This document records the data cleaning steps for individual survey responses in the 2021 American Samoa small boat fishery cost-earning survey data. Each step was manually entered in Excel, with the resulting file saved as "AS PIFSC Data File_v2_manual.csv" in the R project folder path *AmSamoa_SBF > Data*. Each question is listed by question column according to the raw data set. Additional data cleaning steps are described in "AS_SBF_DataCleaning.R" in the *Data* folder.

*Note: Survey hard copies were checked after an initial round of data cleaning. Some steps listed below may no longer apply to the final data set. Please ignore items below that are not reflected in the final data set.

All questions:

- Input median where respondent recorded a range of values

Q2:

- Adjusted responses to reflect Q3-6
- Recoded "PPP_SCAN_0148", "PPP_SCAN_0153", "PPG_DIRECTOR_1095" to "2" to reflect remainder of survey

Q3A:

- Recoded range to midpoint value
- Recoded undefined range to stated value
- Recoded "PPP_SCAN_0148" and "PPP_SCAN_0149" to "99"

Q3B:

- Recoded ranges to correct range of stated value from Q3A
- Input the range from Q3A response

Q4A-Q4F:

- Recoded values so that survey responses =100%

Q4G:

- Recoded "None" to "77"

Q5A:

- Recoded range to midpoint value
- Recoded undefined range to stated value
- Recoded "none" to "0"
- Recoded to "99" for "PPG_DIRECTOR_1102" to make consistent with remaining survey responses

Q5B:

- Recoded to "1" for "PPG_DIRECTOR_1098", reported 0 shore-based trips in 2020
- Recoded to "99" for "PPG_DIRECTOR_1102" to make consistent with remaining survey responses

Q6A-E:

- Recoded "None" to "77"
- Recoded values so that survey responses =100%
- Changed responses to "1" if Q5 is "0"
- Recoded 6A to "99" for "PPP_SCAN_0162," "PPG_DIRECTOR_1097" reported no shore-based trips

Q7A-Q7B:

- Recoded values so that survey responses =100%
- Recoded to all trips taken in local waters for PPP_SCAN_0148 (remainder of survey indicates they took no boat trips in 2020)

Q8:

- Recoded ranges to midpoint values
- Recoded "99" to "1" person

Q9A, 10A, 11A, 12A:

- When respondent gave a stated value, recoded to reflect stated value

Q9B, 10B, 11B, 12B:

- Recoded range to midpoint value
- Recoded undefined range to stated value

Q14A.1-.3:

- Recoded 0133, 0134, 0145, 0164, 0166 to new "mixed motivations" response (8 in manual survey data csv) and "99" for secondary and tertiary

Q14B:

- Recoded "Charter fishing" to "Charter"

Q15A:

- Recoded "1 - when subsistence fishing" to "1"
- Recoded to "5" for PPP_SCAN_0164

Q16A-D:

- Recoded values so that survey responses =100%
- Q16D: recoded to "99" for "PPP_SCAN_0164", remainder of responses show that they sold their catch

Q17:

- If Q17 "no", recoded Q18-21 "no" to "skip"
- If Q17 "yes", recoded Q18-21 "skip" to "no"
- Recoded Q17 to "yes" for PPP_SCAN_0133

Q17-Q21:

- Recoded Q17 "yes" values so that Q20A-Q20D responses =100%
- Removed Q18C (restaurants/stores is a single category on survey); renamed Q18C-G accordingly

Q18F:

- Recoded "Village/Church Functions" to "Church and village"

Q22:

- Recoded "2" to "1" for PPG_DIRECTOR_1096. Boat owner and operator and responded to Q23-31
- Recoded "99" to "2" for crew member who completed parts of the survey, PPP_SCAN_0148

Q23:

- Adjusted response based on if boat owner or owner-operator

Q23-Q30:

- Fill in missing values to "99" when respondent reported owning a boat
- Recoded ranges to midpoint value
- Recoded undefined range to stated number
- Recoded "20+ ya" (over 20 years ago) to "2000"

Q28:

- Recoded "n/a" to "77"

Q29:

- Recoded range to midpoint value
- Recoded "n/a" to "77"

Q30:

- Recoded to most recent year of a range
- Recoded "0" to "77"

Q31-32:

- Recoded "99" and "77" of partial responses to "0"
- Recoded crew responses from "NA"/"77"/"99" to "0" (includes money spent on each item and amount of item used)
- Recoded ranges to midpoint values
- Recoded undefined range to stated number
- Recoded "4 qrts." to estimated cost (respondent skipped boat size and horsepower)
- Recoded "n/a" to "77"
- Recoded "Catch my own" to "0"
- Recoded "varies" to "77"
- Recoded "Cost is inclusive, its all in the same trip. ex. start off trolling end up bottom fishing" to "0"
- Recoded "PPP_SCAN_0142" responses to 77 for Q32

Q33A-L:

- For partial responses, recoded skipped items to "0"
- Recoded full set of 0's to "99"
- Recoded ranges to midpoint values
- For "PPP_SCAN_0161", recoded "9" in 33H to "99"

Q41A-D:

- Recoded blank (missing) responses as "99"

Q43A-I:

- Recoded "Parrot" to "Parrotfish"
- Recoded "Skip" to "Skipjack tuna"
- Recoded "Skipjack" to "Skipjack tuna"
- Recoded "Snapper (red)" to "Red snapper"
- Recoded "Snapper (silver)" to "Silver snapper"
- Recoded "Surgeon" to "Surgeonfish"
- Recoded "Yellow" to "Yellowfin tuna"
- Recoded "Yellowfin" to "Yellowfin tuna"

Q45H:

- If response in 45G, recoded blank to "99"