NOAA FISHERIES

Office of Science and Technology

Marine Recreational Information Program

MRIP Data User Seminar:
Custom Domain Analyses, Part 1

February 22, 2022

John Foster

Overview

- Public-use Datasets (Survey Micro-data)
- Estimation Approach
- Template Programs
- Example Analyses
- General Guidance on Domain Estimation

NOAA FISHERIES

Public-use Datasets

NOAA FISHERIES

Public-use Datasets

- Resource Links
- Key Fields for Estimation
- Trip
- Catch
- Size

NOAA FISHERIES

Resource Links

- MRIP Data Downloads
- Data User Handbook
- Survey Datasets (SAS, CSV): Trip, catch, and size
- Read Me for Datasets and Template Programs (.DOC)
- Dataset Variables (Data Dictionary) (.XLS)

NOAA FISHERIES

Key Fields for Estimation

- Sample Design
  - var_id, psu_id
  - wp_int, wp_catch, wp_size
- Estimation Domain Definition
  - User-selected (e.g., year, st, wave, mode_fx)
- Estimate Variables
  - User-selected (e.g., landing, release)

NOAA FISHERIES

Trip Datasets

- Trip and angler characteristics
- One record for each individual angler-trip intercept (id_code unique identifier)
- Angler groups linked by leader and prt_code fields
- Design fields: var_id, psu_id, wp_int
- Trip mode, area fished, duration, gear, target species
- Access site information
- Limited angler demographics
- Placeholder records for charter boat and headboat mode effort estimates without intercept records

NOAA FISHERIES

Catch Datasets

- Angler-trip level catch data
- One record for each combination of individual angler-trip (id_code) and species caught on trip (sp_code, common)
- Design fields: var_id, psu_id, wp_catch
- Catch in numbers by disposition (e.g., landing, release, tot_cat)
- Landings in weight (wgt_ab1, kg)
- Includes a single record for each angler-trip without catch (missing sp_code, common)

Size Datasets

- Individual fish length and weight measurements by angler-trip (id_code) and species (sp_code, common)
- Design fields: var_id, psu_id, wp_size
- Length values (lngth) are centerline lengths (mm)
- Weight values (wgt) are total weights (kg)
- Indicator fields (lngth_imp, wgt_imp) identify imputed length or weight values (see Section 2.2 for imputation methodology)
- Includes a single record for each angler-trip without catch (missing sp_code, common)

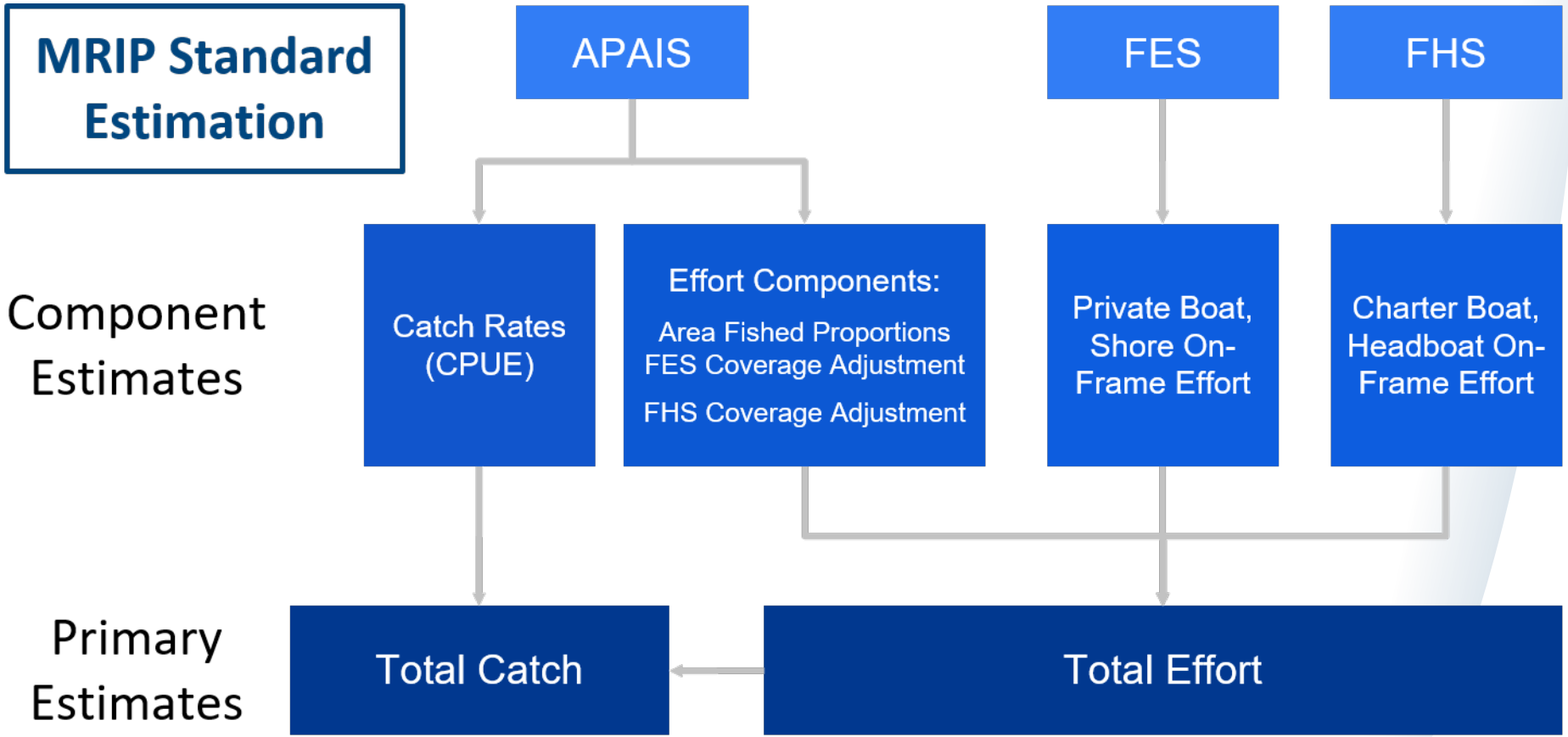
Estimation Approach

NOAA FISHERIES

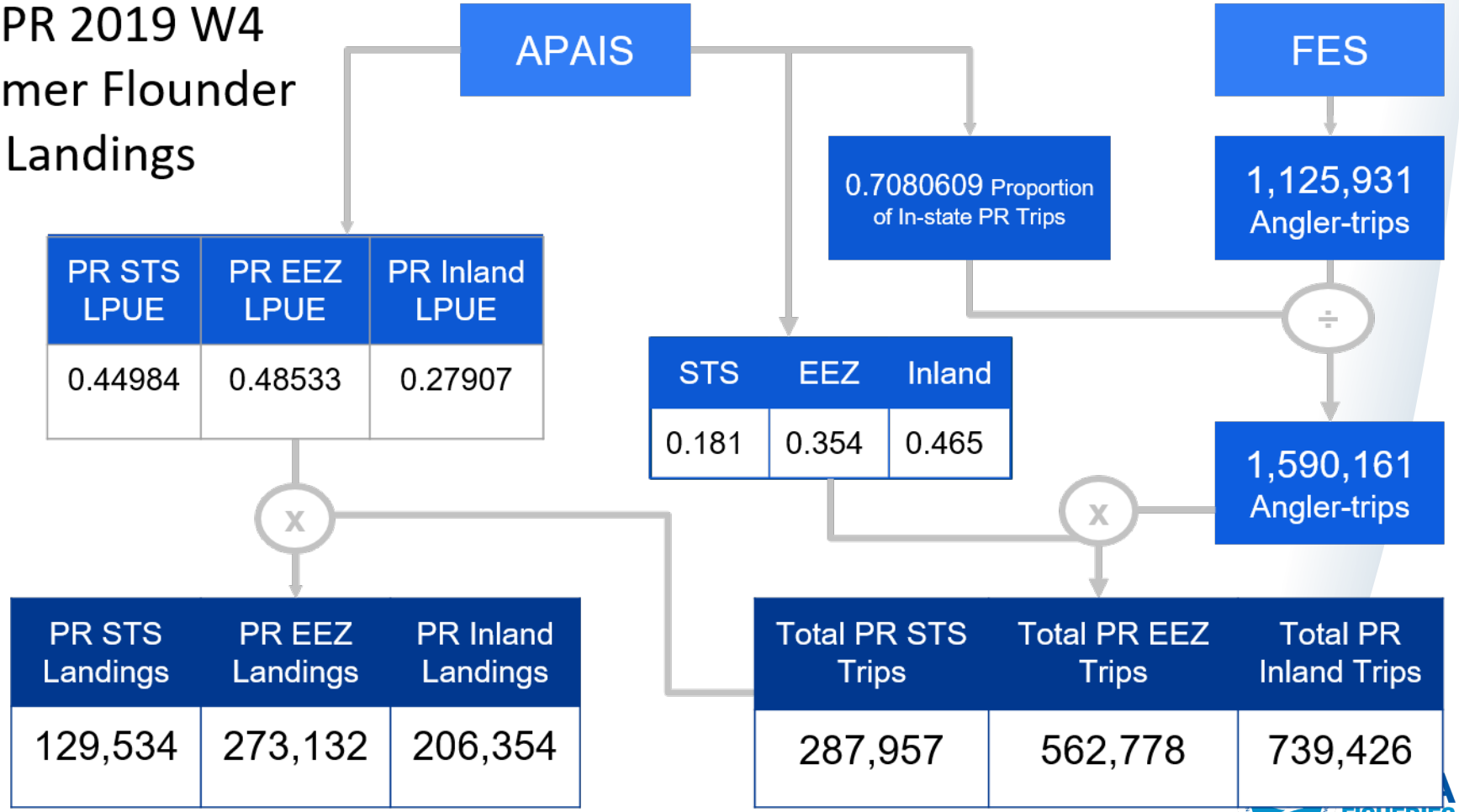
Estimation Approach

- Survey Design and Statistical Methods
  - Section 6.2.3 Total Catch Estimation for Public-Use Datasets
- MRIP Standard Estimation
- Streamlined Estimation for Public-use Datasets
- Important Limitations

MRIP Standard Estimation



NJ PR 2019 W4 Summer Flounder Landings



Streamlined Estimation

- Incorporate the total effort estimates into the APAIS sample weights using a ratio adjustment

- Replace complex MRIP standard estimation with simpler weighted sums to calculate catch and effort estimates

NOAA FISHERIES

Streamlined Estimation

Total Catch = Catch Per Unit Effort X Total Effort

$$\hat{Y}_d = \hat{\bar{y}}_d \times \hat{T}_{T...}$$

$$\hat{Y}_d = \frac{\sum_d w_d y_d}{\sum_d w_d} \times \hat{T}_{T...}$$

$$\hat{Y}_d = \frac{\hat{T}_{T...}}{\sum_d w_d} \times \sum_d w_d y_d = \sum_d \left(\frac{\hat{T}_{T...}}{\sum_d w_d} \right) w_d y_d = \sum_d w_d^* y_d$$

NOAA FISHERIES

Important Limitations

- Streamline estimation approach
  - Point estimates match standard estimates
  - Variances (CVs, PSEs) may not; developing method to address
- Individual catch data at the trip level may have decimal values
  - Complication of having grouped and separated catch among anglers fishing in the same group/party
  - Standard estimation uses different sample weights for grouped catch and separated catch
  - Adjusting catch fields allows use of a single sample weight in streamlined estimation

NOAA FISHERIES

Template Programs

NOAA FISHERIES

Template Programs

- Resource Links
- Analysis Software
- Catch
- Effort (Directed)
- Length Frequencies

Resource Links

- Template Programs
- Read Me for Datasets and Template Programs (.DOC)
- Applied Survey Data Analysis

Analysis Software

- SAS, R, Stata, SPSS, SUDAAN, WesVar
- SAS
  - Survey Procedures
  - Surveymeans
- R
  - Survey Package
- Spreadsheet

NOAA FISHERIES

Catch

- Catch Estimation Template Program
- SAS, R versions
- Uses trip and catch datasets as inputs
- Produces catch estimates as weighted sums or totals within user-defined domains

Effort

- Direct Angler Trip Estimation Template Program
- SAS, R versions
- Uses trip and catch datasets as inputs
- Produces effort estimates as weighted angler-trip sums or totals within user-defined domains

Length Frequencies

- Length Frequencies Template Program
- SAS only
- Uses trip and size datasets as inputs
- Produces both relative length frequencies and absolute length frequencies (catch-at-size) within user-defined domains
- Single species
- Default 2-inch size bins

NOAA FISHERIES

Example Analyses

NOAA FISHERIES

Examples

- SAS
  - Recreate catch estimates available from MRIP queries
  - Custom geographic domain, catch estimates
  - Custom temporal domain, effort estimates
- Spreadsheet example, catch estimates

General Guidance on Domain Estimation

NOAA FISHERIES

General Guidance

- Data User Handbook
    - Section 5.1 General Guidance: Consideration for Domain Analysis
- Domain Definitions and Survey Design
- Sample Sizes
- Precision

Domain Definitions and Survey Design

- Domains are not limited to survey design strata
  - Domains can be smaller, larger, or include subsets of multiple strata
- Available data generally **won't** support very small temporal or spatial domains
  - Individual site estimates
  - Individual day estimates
  - Individual site-day estimates
- Should check for data gaps or levels in domain definitions with no observations

NOAA FISHERIES

Sample Sizes

- Are all levels of defined domains represented in the data?
  - If not, consider collapsing or broadening how domains are defined
  - Conceptually, missing levels are represented by non-missing levels
- No magic number for sample size, more is always better
  - 5 may be sufficient in some cases, 30 may not be sufficient in others
  - How many domain levels are there?
  - How variable are the data (e.g., regulations: 1 vs 25 fish per trip)?
- PROC Surveymeans
  - NOBS will give number of observations
  - NCLUSTER will give number of PSUs

Precision

- Precision is a function of inherent variability of the item being measured and key survey design aspects including sample size
- Estimates calculated from small sample sizes have a high chance of being imprecise
- Defining very small domains increases the likelihood of having small sample sizes and resulting estimates being imprecise
- Always check precision of estimates (CV, PSE)
    - If estimates are imprecise (e.g., PSE>50 or CV>0.5) consider collapsing or broadening domain definitions

NOAA FISHERIES

NOAA FISHERIES

Office of Science and Technology

Marine Recreational Information Program

MRIP Data User Seminar:
Custom Domain Analyses, Part 1

February 22, 2022

John Foster