

# Get and summarize AFSC survey data

Kirstin Holsman, Alaska Fisheries Science Center

## Contents

<b>Overview</b>	<b>1</b>
Step 0: Set up the R workspace . . . . .	3
Step 1: Update SQL queries . . . . .	3
Step 2: Update the LWA regressions . . . . .	4
Step 3: Get CPUE data from the surveys . . . . .	4
Step 4: calc propB for use in biomass weighting age or diet data . . . . .	6
Step 5: Get annual biomass by age and CEATTLE bins: . . . . .	6
Appendix 1: R/setup.Rprimary setup script . . . . .	6

**AFSC Survey CPUE data:** [github.com/kholsman/AFSC\\_CPUE](https://github.com/kholsman/AFSC_CPUE) Repo maintained by:  
Kirstin Holsman  
Alaska Fisheries Science Center  
NOAA Fisheries, Seattle WA  
**kirstin.holsman@noaa.gov**  
*Last updated: Feb 16, 2023*

## Overview

The below scripts return a list object `cpue_data` saved as a compressed Rdata file with the naming 'reg.srvy#.spp.cpue\_data.Rdata' such as "ebs.srvy98.plk.cpue\_data.Rdata". Each `cpue_data` list contains 8 data.frames:

```
load("data/out/2023_01_06/cpue/ebs/ebs.srvy98.plk.cpue_data.Rdata")

names(cpue_data)

## [1] "totalB_N_SEBS_NBS"          "totalB_N_SEBS"
## [3] "mnCPUE_strata_yr"           "total_bin_B_N_SEBS"
## [5] "total_bin_B_N_SEBS_NBS"     "mnCPUE_strata_bin_yr_SEBS"
## [7] "mnCPUE_strata_bin_yr_SEBS_NBS" "CPUE_station_bin_yr"
## [9] "CPUE_station_yr"            "propByBin"
## [11] "propByStrata"               "propByStrataBin_SEBS"
## [13] "propByStrataBin_SEBS_NBS"
```

The data.frames within each `cpue_data` object are

1. **totalB\_N\_SEBS\_NBS**: Total biomass (kg) or abundance (# of fish) for the species in each year for NEBS + SEBS survey results
2. **totalB\_N\_SEBS**: Total biomass (kg) or abundance (# of fish) for the species in each year for only SEBS survey results
3. **mnCPUE\_strata\_yr** : Average survey CPUE (kg per Km2) or abundance (# per Km2) for the species in each strata and year
4. **total\_bin\_B\_N\_SEBS**: Total biomass (kg) or abundance (# of fish) for each bin (10 mm) for the species in each year for NEBS + SEBS survey results
5. **total\_bin\_B\_N\_SEBS\_NBS** : Total biomass (kg) or abundance (# of fish) for each bin (10 mm) for the species in each year for only the SEBS survey results
6. **mnCPUE\_strata\_bin\_yr\_SEBS** : Average survey CPUE (kg per Km2) or abundance (# per Km2) for each size bin for the species in each strata and year in the SEBS
7. **mnCPUE\_strata\_bin\_yr\_SEBS\_NBS** : Average survey CPUE (kg per Km2) or abundance (# per Km2) for each size bin for the species in each strata and year in the SEBS and NEBS
8. **CPUE\_station\_bin\_yr**: Station specific survey CPUE (kg per Km2) or abundance (# per Km2) for each size bin for the species in each year
9. **CPUE\_station\_yr**: Station specific survey CPUE (kg per Km2) or abundance (# per Km2) for the species in each year
10. **propByBin**: proportion of biomass in each size bin per species per year
11. **propByStrata**: proportion of biomass in each strata per species per year
12. **propByStrataBin\_SEBS**: proportion of biomass in each bin and strata per species per year just SEBS
13. **propByStrataBin\_SEBS\_NBS**: proportion of biomass each bin and strata per species per year NEBS + SEBS

These are calculated from the RACEBASE data tables for survey results where total CPUE was recorded for the species  $s$  (location\_catch) at each haul, expanded to include stations  $i$  where CPUE=0 (location) and expanded to each size bin  $l$  using the proportional subset of frequency of fish of given length (mm), binned into 10 mm bins ( $l$ ) and predicted weight ( $\hat{W}$ ) for each size bin  $l$  at each station  $i$ :

$$B_{s,y} = CPUE_{s,k,y} \cdot A_k$$

where  $A_k$  is the area of the strata  $k$  in  $Km^2$  and  $CPUE_{s,k,y}$  is the strata specific average CPUE (kg per  $Km^2$  or number per  $Km^2$ ) of all stations  $i$  in strata  $k$ :

$$CPUE_{s,k,y} = \frac{1}{n_k} \cdot \sum_{n_k} CPUE_{s,k,y,i}$$

where \$ CPUE\_{\{s,k,y,i\}} \$ is the station specific CPUE (saves as the object `cpue_data$CPUE_station_yr`).

To obtain population level estimates of the biomass or abundance of fish by size bin  $l$ , we used a length weight regression to estimate the weight of each size fish  $j$  measured ( $\hat{W}$ ) to calculate the proportion by weight or frequency at each station where

$$\hat{W} = \alpha_s + L_j^{\beta_s}$$

where  $\alpha_s$  and  $\beta_s$  are fit (updated each year) to all available length and weight data from surveys from all years across the EBS, GOA, and AI (rather than each basin separately) and

$$p_{l,i}^w = \frac{N_{l,i} \hat{W}_{l,i}}{\sum N_{l,i} \hat{W}_{l,i}}$$

and

$$p_{l,i}^N = \frac{N_{l,i}}{\sum N_{l,i}}$$

This was then multiplied by the CPUE at each station ( $CPUE_{s,k,y,i}$ ) to obtain a station estimate of CPUE by size bin  $l$

$$CPUE_{s,k,y,l,i} = p_{l,i}^N \cdot CPUE_{s,k,y,i}$$

Finally, the average strata CPUE ( $CPUE_{s,k,y,l}$ ) and whole of EBS biomass by size bin ( $B_{s,y,l}$ ) was calculated as:

$$CPUE_{s,k,y,l} = \frac{1}{n_k} \cdot \sum_{n_k} CPUE_{s,k,y,l,i}$$

and

$$B_{s,y,l} = \frac{CPUE_{s,k,y,l} \cdot A_k}{\sum_{n_k} CPUE_{s,k,y,l} \cdot A_k} \cdot B_{s,y}$$

# Code

.

## Step 0: Set up the R workspace

The first step is to set up the switches for what files to update and create in the file `R/setup.R`. The code below then loads these settings as well as base data, functions, and packages.

```
# get everything set up:
#-----
# rm(list=ls())
# this uses the password saved in R/password.R
suppressMessages(source("R/make.R"))
```

## Step 1: Update SQL queries

This step must be run on a computer that has access to RACEBASE. The code below will generate the base files for steps 2 and 3 below, and will save them in the folder `data/in/2023_02_16` under subfolders for each region in `srvys$reg` and each species in `splist` (see `R/setup.R` to change these settings).

**IMPORTANT:**

- This step must be run in 32 bit R (RODBC doesn't run in 64 bit R) and must be connected to the RACEBASE SQL database
- To change R studio from the default 64 bit to 32 bit go to Tools>Global options and select the 32 bit version of R.
- The code will connect to the SQL database using your password and username. Remember to update the `username_path` in the first line of the `R/setup.R` file and corresponding username and password under `username_password.R`. A template is available under `R/`.

```
# update the SQL queries
#-----
source(file.path(code.path,"R/sub_scripts/runRACE_qrys.R"))
```

## Step 2: Update the LWA regressions

The default code for RACEBASE uses set LW relationships, however we prefer to update the LW regressions using glms. Depending on how many observations exist the LW relationships can be region specific or use data across all regions. The default below is all regions combined. This code generates two outputs in data/out/2023\_02\_16, LWGlms\_srvy\_all\_regs.Rdata and LW\_SmryTable.Rdata. It also updates the species\_lkup\$LW\_a and species\_lkup\$LW\_b' parms used in Step 3.

```
# update the LW regressions
#-----

if(update_LWdata==1){
  source(file.path(code.path,"R/sub_scripts/updateLW.R"))
  # reload with updated data:
  source(file.path(code.path,"R/load_data.R"))
}
species_lkup
```

## Step 3: Get CPUE data from the surveys

This code is the core script for generating the CPUE\_NUMKM2 and CPUE\_BIOMKM2 values by size bin, region, and species.

```
nreg <- length(srvys$reg)
nspp <- length(species_lkup$sp)

for (r in 1:nreg){
  for(s in 1:nspp){
    flnm <- paste0(srvys[r,$reg,".srvy",
                  srvys[r,$num,".",
                  species_lkup[s,$sp]
    cat("now getting data for: ",flnm,"\n")
    cpue_data <- suppressMessages(
      get_CPUE_DATA(
        datapath = data.path,
        out_dir  = file.path(data.out),
        flnm     = flnm,
        species  = species_lkup[s,$SPECIES_CODE,
        survey   = srvys[r,$num,
        includeNBS = TRUE,
        saveit    = T,
        bins     = sp_bins[[ species_lkup[s,$sp ]]])

    ## check the data :
    if(1==10){
      tt <- cpue_data%>%
        group_by(YEAR,REGION,STATIONID,SN)%>%
```

```

        filter(BIN ==400)%>%
        summarize(cnt =length(STATIONID))
        max(tt$cnt) #Should be 1
#this looks to be a duplicate sampling...
#mis-entry or code error ?
        cpue_data%>%filter(YEAR==1988,STATIONID=="J-13")
    }
    rm(cpue_data)
}
}

```

The cpue files are now saved in the directory data/out/2023\_02\_16/..

```

source("R/make.R")
load("data/out/2023_01_06/cpue/ebs/ebs.srvy98.pcod.cpue_data.Rdata")

names(cpue_data)

library(dplyr)
#double check the results
cnt_ByStrataBin <- cpue_data$propByStrataBin_SEBS_NBS%>%
  select("REGION", "YEAR", "STRATUM", BIN, BIN_mm,
         SPECIES_CODE, CN, SN, sp, num,
         "propB_ykl", "propN_ykl")%>%
  group_by(YEAR, REGION, SN, CN)%>%
  summarise(sum_propB_ykl=sum(propB_ykl, na.rm=T),
            sum_propN_ykl=sum(propN_ykl, na.rm=T))

cnt_ByStrataBin_SEBS <- cpue_data$propByStrataBin_SEBS%>%
  select("REGION", "YEAR", "STRATUM", BIN, BIN_mm,
         SPECIES_CODE, CN, SN, sp, num,
         "propB_ykl", "propN_ykl")%>%
  group_by(YEAR, REGION, SN, CN)%>%
  summarise(sum_propB_ykl=sum(propB_ykl, na.rm=T),
            sum_propN_ykl=sum(propN_ykl, na.rm=T))

cnt_ByStrata <- cpue_data$propByStrata%>%
  select("REGION", "YEAR", "STRATUM",
         SPECIES_CODE, CN, SN, sp, num,
         "propB_yk", "propN_yk")%>%
  group_by(YEAR, REGION, SN, CN)%>%
  summarise(sum_propB_yk=sum(propB_yk, na.rm=T),
            sum_propN_yk=sum(propN_yk, na.rm=T))

cnt_ByBin <- cpue_data$propByBin%>%
  select("REGION", "YEAR", BIN, BIN_mm, num,
         SPECIES_CODE, CN, SN, sp,
         "propB_y1", "propN_y1")%>%
  group_by(YEAR, REGION, SN, CN)%>%
  summarise(sum_propB_y1=sum(propB_y1, na.rm=T),
            sum_propN_y1=sum(propN_y1, na.rm=T))

```

**Step 4: calc propB for use in biomass weighting age or diet data**

see above

**Step 5: Get annual biomass by age and CEATTLE bins:**

**Appendix 1: R/setup.Rprimary setup script**