Salmon Salvage NegBin Model

Brian Mahardja

2022-08-24

# Estimating Salvage via Negative Binomial Regression

Authors: Brian Mahardja, Catarina Pien, Michael Beakes

## Background

To evaluate potential changes to the number of length-at-date winter-run and spring-run Chinook Salmon salvaged at the CVP and SWP pumping facilities based on the alternatives, Reclamation analyzed historical salvage data via negative binomial regression. Negative binomial regression requires estimation of a dispersion parameter rather than assuming the variance is equal to the mean. In doing so, negative binomial regression can account for overdispersion, which is common in ecological data (e.g., the salvage dataset) as well as reduce the likelihood of biased coefficient estimation.

monthly\_data <- read.csv("negbinmodel\_monthly\_dataset.csv")  
  
str(monthly\_data)

## 'data.frame': 336 obs. of 14 variables:  
## $ year : int 1993 1993 1993 1993 1993 1993 1993 1993 1993 1993 ...  
## $ month : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ sac\_flow : num 48261 48596 49339 43213 24955 ...  
## $ san\_joaquin\_flow : num 4120 3035 2702 3421 3610 ...  
## $ export : num 11671 9331 6037 5757 3497 ...  
## $ delta\_outflow : num 57886 55022 63969 44296 25188 ...  
## $ omr\_flow\_extrap : num -8014 -6461 -3668 -3202 -836 ...  
## $ winter\_lad\_loss : num 82.55 44.96 4.03 2.01 0 ...  
## $ winter\_expanded\_salvage: num 18.89 12.61 2.58 1.13 0 ...  
## $ spring\_lad\_loss : num 0 0 2.78 186.76 242.57 ...  
## $ spring\_expanded\_salvage: num 0 0 1.68 110.13 139.77 ...  
## $ sac\_trawl\_wr\_count : num 3.875 2.667 6.565 0.143 0 ...  
## $ sac\_trawl\_sample\_size : num 8.62 9.33 9.78 9.76 9.95 ...  
## $ sac\_trawl\_sr\_count : num 0 0 5.13 73.05 2.9 ...

## Methods

### Data

Winter-run and spring-run length-at-date Chinook Salmon salvage records from January 1st 1993 to December 31st 2020 were gathered from the California Department of Fish and Wildlife salvage database posted at the SacPAS website (<http://www.cbr.washington.edu/sacramento/data/query_loss_detail.html>). To incorporate hydrodynamic effects on salvage count into the models, Delta export (QEXPORT), Sacramento River flow (QSAC), and San Joaquin River (QSJR) were extracted from the California Department of Water Resources Dayflow data (<https://data.cnra.ca.gov/dataset/dayflow>). Additionally, combined Old and Middle River flow data were pulled from the U.S. Geological Survey National Water Information System website (<https://nwis.waterdata.usgs.gov/nwis>; stations 11313405 and 11312676). Because data gaps exist in the Old and Middle River flow data, ordinary least squares regressions were conducted so that each dataset can be used to predict, and therefore fill, missing data in the other (adjusted : 0.97). Lastly, to account for the variable numbers of juvenile Chinook Salmon entering the Delta by year and month, Sacramento Trawl data were acquired from the Delta Juvenile Fish Monitoring Program through the ‘deltafish’ package available on GitHub (<https://github.com/jeanetteclark/deltaFish>). Sacramento Trawl catch per unit trawl for each day was calculated for both winter-run sized and spring-run sized Chinook Salmon.

Monthly average value summary table

| month | mean\_winter\_lad\_loss | sd\_winter\_lad\_loss | mean\_spring\_lad\_loss | sd\_spring\_lad\_loss |
| --- | --- | --- | --- | --- |
| 1 | 16.4082604 | 27.1491258 | 0.1326613 | 0.3537573 |
| 2 | 24.5842804 | 41.2832423 | 0.9332653 | 1.8764361 |
| 3 | 45.2405876 | 86.5127744 | 78.8869585 | 121.2230005 |
| 4 | 4.9828452 | 9.4312297 | 477.1068452 | 674.3915428 |
| 5 | 0.1888940 | 0.4313571 | 231.7682143 | 366.1717628 |
| 6 | 0.0092976 | 0.0491984 | 25.7662262 | 67.2633749 |
| 7 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 8 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 9 | 0.0000000 | 0.0000000 | 0.0310000 | 0.1640366 |
| 10 | 0.0000000 | 0.0000000 | 0.0453456 | 0.2399465 |
| 11 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 12 | 7.8927304 | 14.4122838 | 0.0000000 | 0.0000000 |

For each variable, data were averaged by month and year with missing data removed. Because monthly salvage values tend to be low or mostly zeroes in for most months out of the year, only December to April period was used for winter-run Chinook Salmon analysis, and only March to June period was used for spring-run Chinook Salmon analysis. Overdispersion was apparent during initial inspection the response variable data (mean variance) supporting the use of negative binomial regression in this analysis.

