Sample size needed to approximate weight-frequency distributions of Togiak herring

Jane Sullivan

February 2019

Recipients: Sherri Dressel and Sara Miller

Background

The objective of this analysis is to determine how many samples are needed in order to approximate the underlying weight-frequency distribution of a population. The managers use this daily distribution of weights to decide when to close the fishery.

Methods

We used 2018 weight data (g) collected across the entire fishery season. Exploratory analysis showed two samples with weights of 0 grams. These were assumed to be N/A values, and were removed from the remainder of the analysis.

A plot of the cleaned dataset is shown below:
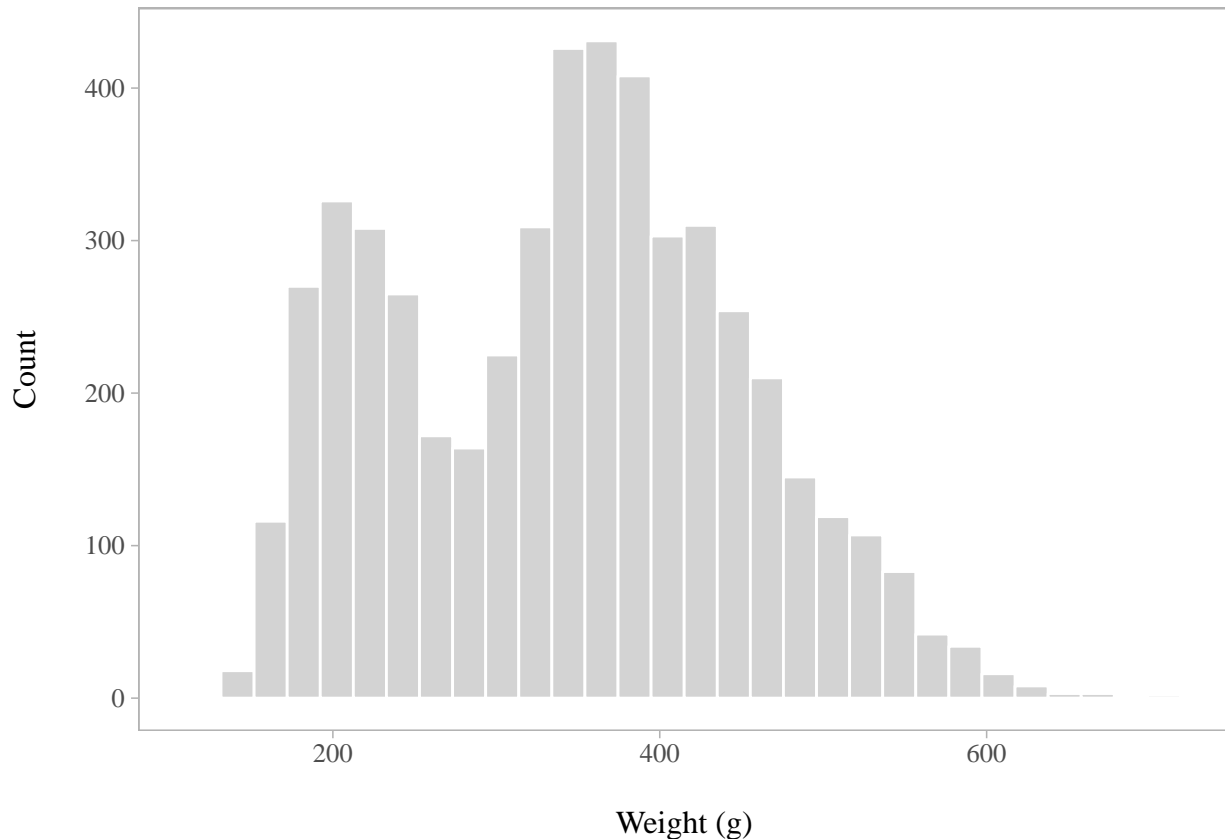
The data show a bimodal distribution, with modes at rougly 200 and 375 g. We were interested in using a bimodal distribution to set sampling goals, because it is likely the most complex distribution that biologists will encounter in-season.

We defined the reference population distribution by resampling the original data 10,000 times with replacement. We used three performance metrics to set sampling goals:

1) mean weight,
2) 25-gram bins (which we call our fine-scale resolution), and
3) 50-gram bins (our coarse resolution).

Plots of the reference population under the two binning schemes are shown below:
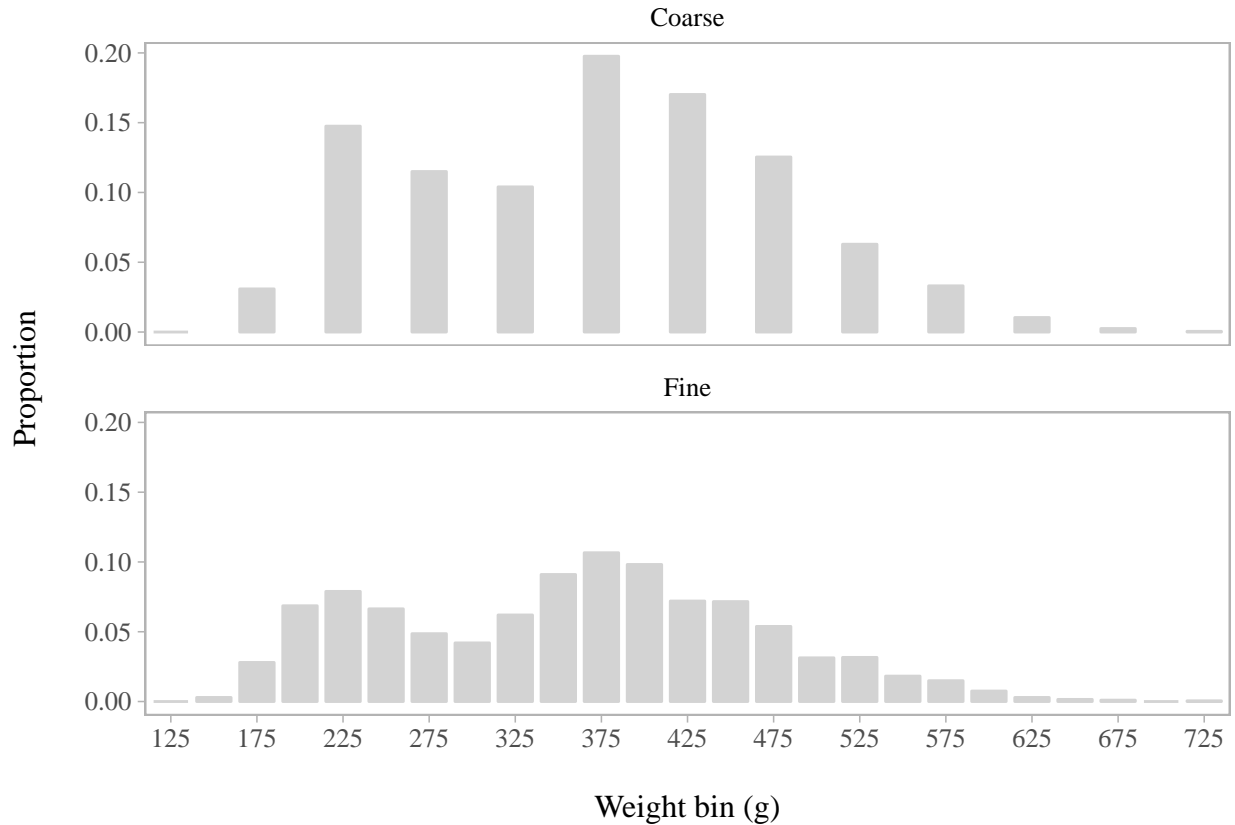
Following Miranda (2007), we took random samples without replacement of 12 sample sizes (n), where n included 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1,000. We took each of these sample sizes 1,000 times without replacement from the reference population, and the results were compared with those of the reference population.

Each of the 12 sets of 1,000 samples were evaluated using the 3 performance metrics stated above. For mean weight, equivalence to the reference population was assessed by computing the fraction of the 1,000 estimates that fell within 10% (5% on either side) of the reference weight. For the 25- and 50-gram weight-frequency distributions, equivalence was assessed by computing the fraction of the 1,000 weight-frequency distributions that differed by 10% or less from the reference weight-frequency distribution, as measured with Renkonen's percentage similarity index (PS) (Krebs 1999), where

$$PS = \sum_{i=1} \min(p_{1,i}, p_{2,i}),$$

2

where $p_{1,i}$ is the proportion of the ith bin in the reference population and $p_{2,i}$ is the proportion of the ith bin in the sample.

Results

The figure below shows the probability of estimating the weight-frequency within 10% of the true value based on the chosen performance metrics.

The ability to estimate the mean of the distribution requires the smallest sample size. It requires additional samples in order to approximate the distribution using a coarse resolution binning approach of 50-g bins, and even more to approximate the fine scale resultion bins.
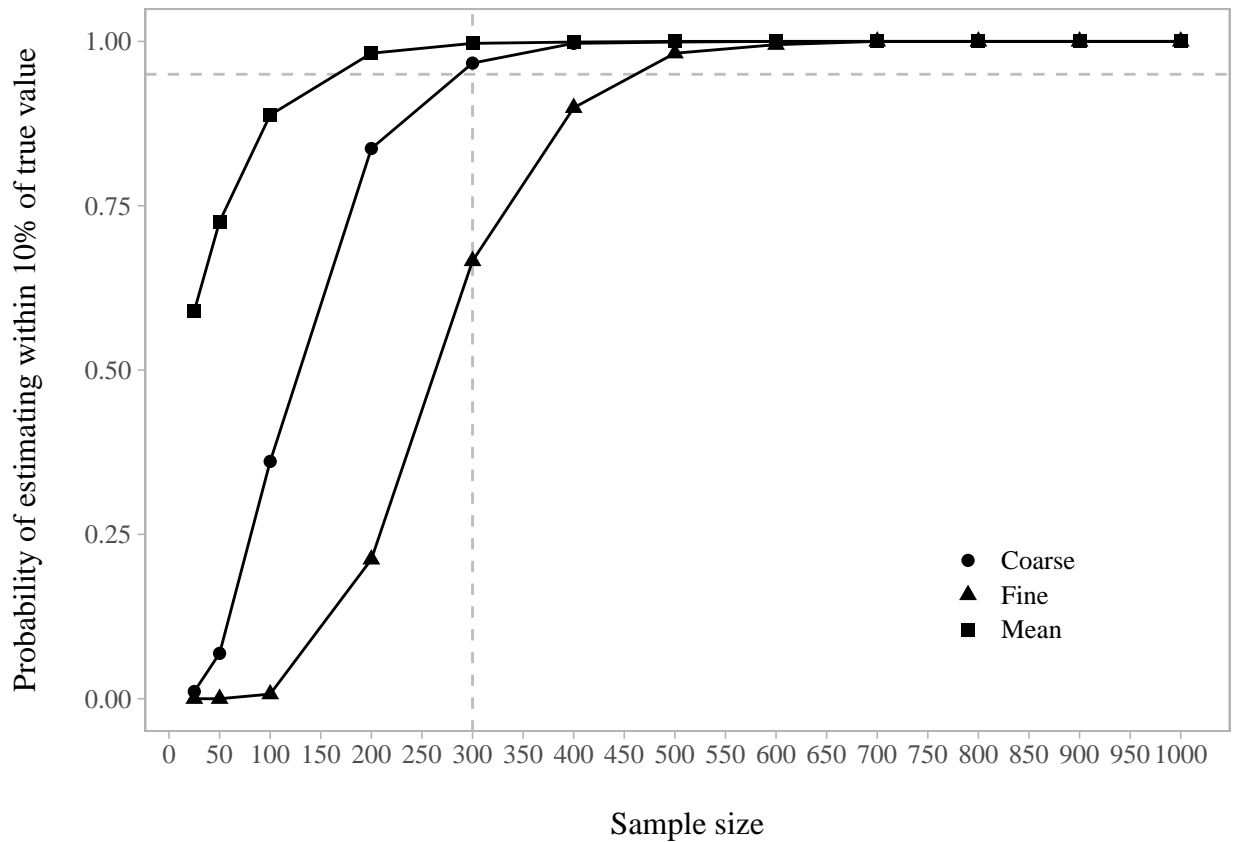
Recommendation/Discussion

Based on this analysis, a sample size of 300 is needed in order to approximate the weight-frequency distribution 95% of the time using a coarse resolution binning of 50-g.

If resources were insufficient to reach a sample size of 300, there are two paths forward. First, one could increase the limit from 10% within the true value to 15 or 20% and re-calculate the sample size needed to satisfy this limit. The selected 10% limit was based on Miranda (2007) and will provide robust estimates for in-season management. However, if most of the true distributions are unimodal instead of bimodal, a smaller sample size will likely be sufficient in order to approximate it. Further analysis would be needed to confirm this. Alternatively, one could accept a lower probability of estimating the true value within 10%. For example, a sample size of 250 (252 to be exact) will result in a 90% probability of estimating the true

value within 10% using 50-g bins. Similarly, a sample size of 200 results in an 84% probability of estimating the true value within 10% using 50-g bins.

Although reducing bin size resolution increases the probability of approximating the distribution under smaller sample sizes, we do not recommend reducing the resolution below 50-g because it could mask modes in the distribution.

References

Krebs, C. J. 1999. Ecological methodology, 2nd edition. Addison-Welsey Educational Publishers, Menlo Park, California.

Miranda, L. E. 2007. Approximate sample sizes required to estimate length distributions. Transactions of the American Fisheries Society 136:409–415.

R Session Info

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] bindrcpp_0.2.2  knitr_1.20      forcats_0.3.0   stringr_1.3.1
##  [5] dplyr_0.7.6     purrr_0.2.5     readr_1.1.1     tidyr_0.8.1
##  [9] tibble_1.4.2    ggplot2_3.0.0   tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.17     cellranger_1.1.0 pillar_1.2.3     compiler_3.5.1
##  [5] plyr_1.8.4       bindr_0.1.1      tools_3.5.1      digest_0.6.15
##  [9] lubridate_1.7.4  jsonlite_1.5     evaluate_0.10.1  nlme_3.1-137
## [13] gtable_0.2.0     lattice_0.20-35  pkgconfig_2.0.1  rlang_0.2.1
## [17] psych_1.8.4      cli_1.0.0        rstudioapi_0.7   yaml_2.1.19
## [21] parallel_3.5.1   haven_1.1.2      withr_2.1.2      xml2_1.2.0
## [25] httr_1.3.1       hms_0.4.2        rprojroot_1.3-2  grid_3.5.1
## [29] tidyselect_0.2.4 glue_1.2.0       R6_2.2.2         readxl_1.1.0
## [33] foreign_0.8-70   rmarkdown_1.10   modelr_0.1.2     reshape2_1.4.3
## [37] magrittr_1.5     backports_1.1.2  scales_0.5.0     htmltools_0.3.6
## [41] rvest_0.3.2      assertthat_0.2.0 mnormt_1.5-5     colorspace_1.3-2
## [45] labeling_0.3     stringi_1.1.7    lazyeval_0.2.1   munsell_0.5.0
## [49] broom_0.4.5      crayon_1.3.4
```