

Simulation Study to Evaluate Alternative Estimators of the Mean Avidity of Marine Recreational Fishing Participants from Access Point Survey Data

FY 2014 Proposal

Dave Van Voorhees
Created: 05/13/2015

1. Overview

1.1. Sponsor

Ron Salz

1.2. Focus Group

Survey Design and Evaluation

1.3. Background

The Marine Recreational Fisheries Statistics Survey (MRFSS) estimator of the number of anglers participating in marine recreational fishing (participation) is based on the Coastal Household Telephone Survey (CHTS) estimator of total fishing effort and a secondary Access Point Angler Intercept Survey (APAIS) estimator of angler avidity. The number of participants is estimated by dividing the APAIS estimate of angler avidity (mean number of fishing days per participant) into the MRFSS estimate of total fishing effort (total number of angler fishing days). The National Research Council's Review of NOAA's recreational fishery survey methods pointed out that the MRFSS APAIS estimators assumed simple random sampling and did not properly weight data to take the complex, stratified, multi-stage cluster sampling design of the APAIS into account. A project team was established in 2011 to develop an appropriately weighted APAIS estimator of angler avidity and evaluate possible bias in the MRFSS estimators of angler avidity and total participation. The MRFSS APAIS estimator of angler avidity was the harmonic mean of the number of days on which interviewed fishing participants reported they had fished during the last two or twelve months. To remove potential for bias in this estimator, the project team worked to develop a properly weighted version that could potentially be used both to evaluate bias in the MRFSS estimators and to calculate more accurate historical estimates of marine recreational fishing participation. However, an analysis of the weighted harmonic mean of reported days fished showed that it is not a design-unbiased estimator of angler avidity under the complex APAIS sampling design. The weighted estimator is likely to be much less biased than the MRFSS unweighted estimator, but the team determined it would be useful to measure the extent of any remaining bias in the weighted estimator. The project team proposes to implement a statistical simulation study to enable us to better understand the performance of the weighted harmonic mean as an estimator of angler avidity.

1.4. Project Description

The first step is to closely investigate the characteristics of angler populations using past MRFSS data. Because the CHTS utilizes a random digit dialing phone survey design, household residents who take very few marine recreational fishing trips should be just as likely as those who take many fishing trips to be interviewed through this sampling approach. On the other hand, the sampling design of the APAIS is much more likely to be subject to an "avidity bias". Anglers who take a lot of fishing trips will have a greater chance of being intercepted and interviewed than those who take very few trips. Thus, we plan to focus on the past CHTS data to study the characteristics of typical angler populations, such as the mean and variance of fishing avidity during a given two month period. The results will be used as the baseline characteristics to define parameter variables and their values to generate various simulated angler populations. Once the baseline characteristics of a typical angler population are defined, simulations will be carried out under various conditions to generate sets of angler populations with different characteristic with respect to how anglers with different levels of fishing avidity are distributed among fishing sites that differ in fishing activity. For example, a simulated angler population might be distributed among sites and time intervals under the assumption that there is no consistent correlation between the expected fishing pressure and the avidity of anglers fishing within a given site and time interval (i.e., anglers with different level of avidity are randomly assigned to fishing sites and time intervals with different levels of fishing pressure). Alternatively, a simulated population might be distributed under the assumption that angler avidity is related to the expected fishing pressure (i.e., anglers with high avidity are more often found fishing at sites and times with low fishing pressure than at sites and times with high fishing pressure). Various probability sampling designs, including that used by the APAIS, will be applied to the simulated populations. Estimates of angler avidity and estimated standard errors of those estimates will be calculated using a weighted harmonic mean from the samples drawn iteratively from simulated angler populations. This will allow evaluation of differences between point estimates of angler avidity and the true value in each simulated population. In addition, it will be possible to evaluate the true variance of the point estimator over many independent iterations of sampling. The results will be analyzed to compare and evaluate the robustness of the estimator under different population distribution assumptions. It should be possible to determine the direction and magnitude of any consistent bias in the weighted estimator of angler avidity. This information will prove useful in evaluating the total extent of bias in the MRFSS estimators of angler avidity and participation and in calculating more accurate historical estimates of participation.

1.5. Public Description

1.6. Objectives

The primary objective of this effort is to evaluate the relative robustness of both the unweighted MRFSS APAIS estimator and a new weighted APAIS estimator of angler avidity. The results will be used to better evaluate possible sources of bias in the MRFSS estimator of angler avidity, as well as the performance of the alternative estimator. Identification of a properly weighted APAIS estimator that shows very little bias under a wide variety of conditions is an important objective. Such an estimator could be used to produce revised participation estimates for prior years and allow a good evaluation of the magnitude and direction of

any bias in the past MRFSS estimates that relied on the unweighted MRFSS APAIS estimator of angler avidity.

1.7. References

NRC. 2006. Review of recreational fisheries survey methods. The National Academies Press, Washington, D.C. 'Some comments on the use of the harmonic mean' by Jean Opsomer.

2. Methodology

2.1. Methodology

Simulation study

2.2. Region

Gulf of Mexico, Mid-Atlantic, North Atlantic, South Atlantic

2.3. Geographic Coverage

NA

2.4. Temporal Coverage

NA

2.5. Frequency

NA

2.6. Unit of Analysis

angler

2.7. Collection Mode

NA

3. Communication

3.1. Internal Communication

Monthly conference call meeting and exchange or distribution of materials by email as necessary.

3.2. External Communication

Monthly progress report to the MRIP Operations Team.

4. Assumptions/Constraints

4.1. New Data Collection

N

4.2. Is funding needed for this project?

Y

4.3. Funding Vehicle

New contract needed

4.4. Data Resources

MRFSS APAIS and CHTS data

4.5. Other Resources

Survey research firms, academic consultants and/or their graduate students

4.6. Regulations

None

4.7. Other

None

5. Final Deliverables**5.1. Additional Reports**

Final report summarizing the results

5.2. New Data Set(s)

None

5.3. New System(s)

None

6. Project Leadership**6.1. Project Leader and Members**

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Richard	Aiken		Team Member	US Fish and Wildlife Service			
Mike	Brick		Team Member	Westat			
Han-Lin	Lai		Team Member	NOAA Fisheries			
Jean	Opsomer		Team Member	Colorado State University			
Jun	Rossetti		Team Member	ICF International			
Dave	Van Voorhees		Team Leader	NOAA Fisheries			

7. Project Estimates**7.1. Project Schedule**

Task #	Schedule Description	Prerequisite	Schedule Start Date	Schedule Finish Date	Milestone
1	hire consultant		03/01/2014	04/30/2014	
4	final report	1,2,3	09/01/2014	12/31/2014	
2	kick-off meeting	1	05/01/2014	05/31/2014	
3	simulation study	1,2	06/01/2014	08/31/2014	

7.2. Cost Estimates

Cost Name	Cost Description	Cost Amount	Date Needed
consultant support		\$45000.00	05/01/2014
TOTAL COST		\$45000.00	

8. Risk

8.1. Project Risk

Risk Description	Risk Impact	Risk Probability	Risk Mitigation Approach
Project unable to find suitable consultant/graduate students support in a timely manner	Delay project completion	Low	Initiate the search for appropriate consultant support early in the project schedule
Simulation and analysis proves to be more complex than expected	Delay project completion	Medium	Hire experts in the field to provide technical consulting

9. Supporting Documents

"Harmonic Mean", page 1

Some comments on the use of the harmonic mean

Jean D. Opsomer
Colorado State University

December 24, 2010

Harmonic mean

Currently, MRFSS uses the harmonic mean to estimate the mean number of trips taken by recreational anglers, often referred to as the mean angler *avidity*. This estimator is justified based on the equivalence between the Hajek estimator, which is guaranteed to be an approximately design unbiased estimator, and the harmonic mean, when sampling is done in proportion to the individual angler's avidity.

The equivalence between the Hajek estimator and the harmonic mean is readily seen from the following reasoning. Letting Y_k denote the total number of trips taken by angler k and π_k the corresponding selection probability, the Hajek estimator is defined as

$$\hat{Y}_{HA} = \frac{\sum_s \frac{Y_k}{\pi_k}}{\sum_s \frac{1}{\pi_k}}.$$

Assuming that the sample inclusion probability for angler k is $\pi_k = c Y_k$ for some proportionality constant c , we obtain

$$\hat{Y}_{HA} = \frac{\sum_s \frac{Y_k}{c Y_k}}{\sum_s \frac{1}{c Y_k}} = \frac{\sum_s \frac{Y_k}{c Y_k}}{\sum_s \frac{1}{c Y_k}} = \frac{n}{\sum_s \frac{1}{Y_k}},$$

with the final ratio in this expression equal to the harmonic mean. The inclusion probabilities π_k in these expressions are actually unknown, since to $c = n / \sum_U Y_k$. However, since c cancels in the ratio, this is not a problem when computing this estimator.

Investigation of the suitability of the harmonic mean

The above equivalence between the Hajek estimator and the harmonic mean is lost with more complex designs, even when sampling continues to be proportional to the overall avidity of the individual anglers. For a simple example, consider the case with two strata with sample sizes n_1 and n_2 . In that case, the Hajek estimator is defined as

$$\begin{aligned}\hat{Y}_{HA} &= \frac{\sum_{s_1} \frac{Y_k}{\pi_k} + \sum_{s_2} \frac{Y_k}{\pi_k}}{\sum_{s_1} \frac{1}{\pi_k} + \sum_{s_2} \frac{1}{\pi_k}} \\ &= \frac{\sum_{s_1} \frac{Y_k}{c_1 Y_k} + \sum_{s_2} \frac{Y_k}{c_2 Y_k}}{\sum_{s_1} \frac{1}{c_1 Y_k} + \sum_{s_2} \frac{1}{c_2 Y_k}} \\ &= \frac{\sum_{s_1} \frac{1}{c_1} + \sum_{s_2} \frac{1}{c_2}}{\sum_{s_1} \frac{1}{c_1 Y_k} + \sum_{s_2} \frac{1}{c_2 Y_k}},\end{aligned}$$

with $c_1 = n_1 / \sum_{U_1} Y_k$ and similarly for c_2 . The only way to make this equivalent to the harmonic mean is if $c_1 = c_2 = (n_1 + n_2) / (\sum_{U_1} Y_k + \sum_{U_2} Y_k)$, or equivalently, that the sample allocation to the strata is proportional the unknown total avidity in the strata: $n_1/n = \sum_{U_1} Y_k / \sum_U Y_k$ (and equivalent expression for n_2). Clearly, this is a very strong assumption that is unlikely to be more than approximately met, and it removes all flexibility in sample allocation across strata. A more serious problem is that c_1 and c_2 are unknown, so that even the statistically appropriate Hajek estimator is actually unfeasible.

We now consider the more realistic case with stratification and clustering, which corresponds to the MRFSS design where the cluster is the site and the stratum is a combination of wave, mode and state. In that case, the Hajek estimator is

$$\hat{Y}_{HA} = \frac{\sum_{h=1}^H \sum_{s_{1h}} \frac{1}{\pi_{I_i}} \sum_{s_i} \frac{Y_k}{\pi_{k|i}}}{\sum_{h=1}^H \sum_{s_{1h}} \frac{1}{\pi_{I_i}} \sum_{s_i} \frac{1}{\pi_{k|i}}},$$

where π_{I_i} is the inclusion probability for cluster i and s_{1h} denotes the sample of clusters in stratum h . We assume that the sampling of anglers within the clusters is proportional the anglers total avidity, so that the within-cluster inclusion probability is $\pi_{k|i} = c_i Y_k$, and $c_i = n_i / \sum_{U_i} Y_k$ with n_i the sample size in cluster i and U_i the unknown population total of the avidities for anglers in cluster i . If the c_i are made to be constant, so that the sample size for each cluster is proportional the total avidity in the cluster, then this estimator is equivalent to a weighted harmonic mean, with weights equal to $1/\pi_{I_i}$.

The same problem as above occurs, which is that this is not a realistic nor feasible sample allocation. As above, the Hajek estimator is also not feasible, because it relies on unknown sampling probabilities.

The above discussion focused on the case where sampling is proportional to the anglers' total avidity Y_k , even when applied within strata and clusters. This requires the assumption that anglers can be considered "homogeneous" with respect to their fishing preferences. In the case of clusters, it assumes that the site-specific avidity for an angler, say Y_{ik} for angler k in site i , can be written as $Y_{ik} = a_i Y_k$, i.e. relative preference for site i does not depend on the angler k . If that assumption is reasonable, then sampling proportionally to either Y_k or Y_{ik} is equivalent, and the above discussion continues to apply.

Finally, consider the case where the site-specific avidity Y_{ik} is not proportional to the overall avidity Y_k , in the sense that $Y_{ik} = a_i Y_k$ fails to hold across anglers and sites. In this case, because sampling is not proportional to the variable being measured in the survey, the Hajek estimator does not reduce to the harmonic mean even in the simplest case first considered above, so that the harmonic mean is not design-unbiased.

Recommendations

I was asked to investigate the suitability of the (weighted) harmonic mean to estimate mean avidity based on the APAIS. One issue of concern that was identified previously was the use of total (annual) avidity instead site or stratum specific avidity. That is indeed an issue, as noted at the end of the above discussion. However, the problems in estimating avidity based on the APAIS are in fact even more fundamental.

First, the use of the harmonic mean, whether weighted or unweighted, relies critically on assumptions that allow cancellation of proportionality constants across clusters and strata. These assumptions are unlikely to be met for the APAIS overall and, if they are not met, the harmonic mean is no longer design-unbiased. Whether the assumptions hold, exactly or even only approximately, cannot be checked based on the observed data, since they are based on unobserved total avidity within strata and clusters.

Second, in intercept surveys where the harmonic mean is inappropriate because it is not design-unbiased, more general design-based estimators of mean avidity such as the

Hajek estimator cannot be used either. This is due to the fact that the true sampling probabilities are based on unobserved avidities.

In summary, I would recommend investigating alternative ways to estimate the mean angler avidity that do not rely on the APAIS. The CHTS or other future surveys that do not rely on site intercepts are likely to provide more suitable avenues for obtaining avidity estimates.