Homework 1

Jonah Bacon

FISH 621

01 Feb 2022

**Problem 1:**

In the sample size of 50, the mean length was 39.96 mm with a variance of 353. Converting the variance in mean length to a relative scale results in a CV of 0.47. The variance in our mean length estimate (y\_bar) was 6.71.

For the sample size of 250, the calculated mean length was longer (44.46 mm) with a larger variance as well (638). The standardized variability estimate CV was 0.57, indicating that the data was more variable than our sample with 50 individuals. Finally, the estimated variance in the mean estimate was much lower (1.91) owing to the fact that our sample size was larger.

The CV does not scale with sample size. Instead, the sample CV is a property merely of that particular sample and sample size. On the other hand, the unbiased estimate of the variance of the estimate DOES scale with sample size. Increasing the sample size decreases the estimated variance in y-bar.

**Problem 2:**

From our sample, the average number of flower per quadrat was 14.133 with a variance of 74.84. The CV was 0.612 and the estimated variance in the calculated sample mean was 4.98. The estimated variance in the calculated sample mean estimate was relatively large as compared to the actual mean estimate (4.98 versus 14.133). This may indicate we should have low confidence in our mean estimate.

Scaling the estimated mean number of flower per quadrat up to the size of the field allowed us to estimate the total population size. Our population estimate was 84,800 flowers in the field with a variance of 179162400. The CV for the population estimate total was 0.158. The total sample space (N) has a large effect on the variance of the estimate total because of the fact that it is squared (N^2) in that respective variance equation. The CV ended up being relatively low for the estimate of the total population size, possibly indicating that we can have confidence in our total population estimate even though our mean estimate had a high proportional amount of variance.

**Problem 3:**

The estimated N\_hat (population size) from the three sampling events was 80, 147, and 127 for the first, second, and third sampling events, respectively. The plot visualizes the distributions resulting from the three sampling events. Event 1 resulted in a distribution with relatively lower variation centered on a lower N\_hat estimate. Event 2 had the highest N\_hat estimate and the highest variation around that estimate. Event 3 was intermediate between the two in terms of N\_hat estimate and variation. From visualizing the three PDF’s, it can be seen that they overlap significantly in their distribution tails. This indicates that you might take caution in inferring that the vole population has changed between the three sampling events. Instead, you might reason that the population has been well classified by your samples and that the population size is within the three population size estimates.

**Problem 4:**

The expected number of marked individuals you would observe under the first scenario is 5, which is confirmed by summing the possible m2 values multiplied by their respective probability of occurrence.

Altering the number of whales inspected for marks (n2) results in different distributions for the number of marked individuals we expect to observe. Under a low t2 sampling effort (n2 = 50), our m2 distribution is shifted to the left with an m2=2.5 having the maximum likelihood. Under this scenario, an m2=0 has a probability of 0.072, which is pretty high in my opinion. Not observing any marked individuals at t2 (m2 = 0) would certainly cause a researcher headaches attempting to estimate N\_hat. Under a high t2 sampling effort (n2 = 200), the expected m2 increases to 10 and there is a low probability of observing very few or zero marked individuals. With an intermediate t2 sampling effort (n2 = 100), the expected m2 is unsurprisingly intermediate (E(m2) = 5) and there is a decent chance that very few marked individuals may be observed. This example shows that it is important to put enough effort into the t2 sampling period in order to get a reasonably accurate population size estimate.

If the true population size (N) changes, the m2 PDF’s are similar. Under the scenario of a high population size (N=2000), E(m2) is 2.5 with a high likelihood of m2=0 occurrence (probability = 0.072). With a low population size (N=500), E(m2) is 10 and there is a low probability of observing 0 marked individuals at t2. Based upon these results, it shows the importance of having some sort of preconceived population size estimate prior to designing your study. If the true population size is much larger than you expected, you have a high probability of observing few or no marked individuals at t2. On the other hand, if the true population size is much smaller than anticipated, you may expend much more resources on your t2 sampling effort than you needed to in order to estimate the population size.

**Problem 5:**

There are two tables in my R script (named “chapman\_table” and “bailey\_table”) and at the end of this document that detail the N\_hat estimates, associated variance, and CV for each month according to which model was used.

From the results of this experiment, it is apparent that the population is declining over the course of the season. The distributions are not only shifting downwards in their N\_hat estimates but they are also becoming more precise, which can be visualized by the narrowing of the distribution tails and the shrinking of the density within those tails as the months progress through the year. These results indicate that the population is at risk of being over exploited and that management action should be taken to curb fishing pressure in order to reduce the current seasonal population decline.

The Peterson simple mark-recapture experiment assumes that 1) this is a closed population, 2) all individuals have the same probability of being caught, 3) marking doesn’t affect catchability, 4) there are no major mortality events, and 5) there is no tag loss. In regards to assumption 1, the study most likely does not meet this assumption. You are working on a small river system, which is most likely involved in a much greater watershed. There is high probability that immigration could occur into this system. In regards to assumption 2, fish might learn to avoid being caught throughout the summer or only the “dumb” fish which were aggressive were caught early in the year. Thus “smart” fish or fish caught-and-released remain in the river late in the summer and have a lower chance of catch. In regards to assumption 3, marking involves catching fish. This might lead to a violation of assumption 2 as those fish are less likely to be caught. But I would assume that the actual tag doesn’t affect catchability, but it would be good to verify this with a controlled experiment. In regards to assumption 4, it would be good to monitor the river throughout the season to ensure that there wasn’t a mortality event. The observation that the population estimate decreased consistently throughout the season causes me to infer that it is most likely that a major mortality event did not occur. In regards to assumption 5, it is really hard to determine if there is any tag loss occurring without some sort of prior knowledge. It is quite possible that there could be loss occurring, and a laboratory-based study should be referenced or implemented to gauge that probability.

Chapman table

N\_hat var CV

May 398.30000 4417.01857 0.16686079

June 301.00000 2976.85714 0.18126439

July 283.43750 2483.59949 0.17582615

August 102.50000 34.96622 0.05768999

September 99.72093 26.83052 0.05194315

Bailey table

N\_hat var CV

May 393.25000 6146.87202 0.19936933

June 290.38462 5504.54353 0.25549773

July 277.50000 3876.83824 0.22437572

August 101.25000 153.92736 0.12253577

September 99.06977 87.96893 0.09467243

**TIME ALLOCATION:** 6 hours