Satoshi Ido
3488706
26 February 2023

STAT 522 HW5

1.

(b) Using the sample, estimate the average area for objects in the bin. Give a 95% CI.

The mean is 28.245
The 95% CI of the estimate is [26.41297, 30.07703].

a. How does the 95% CI for this sample compare to the 95% CI from your SRS sample results in Homework 3?

The CI for this sample is underestimated than that of the SRS sample from homework 3.

(c) Using the sample, estimate the total number of gray objects in the population, along with the 95% CI.

The total number of gray objects is 7200, and the 95% CI is [5919.405, 8480.595].

a. How does the 95% CI for this sample compare to the 95% CI from your SRS sample results in Homework 3?

The CI for this sample is also underestimated than that of the SRS sample from HW3. However, the stratification does not have much smaller variance, meaning the stratification method is not necessarily effective.

2.

(c) For each sample, estimate the proportion of players who are pitchers and give a 95% CI.

The SRS estimation of the proportion for a pitcher is 0.520 and the 95% CI is [0.44713114, 0.59286886] (=0.14573772).

The stratified estimation of the proportion for a pitcher is 0.448683, and the 95% CI is [0.373065636, 0.52429948] (=0.151233844).

a. How do the estimates compare across the two samples? How do the 95% CI's compare?

It is clear that the estimate with stratification adjusts the overestimation of SRS estimate. However, the stratification does not have much smaller variance (0.145 vs 0.151), meaning the stratification method is not necessarily effective.

(d) For each sample, estimate the mean of the `logsal` variable and give a 95% CI.

a. How do the estimates compare across the two samples? How do the 95% CI's compare?

The estimates with SRS of `logsal` is 13.788 and 95% CI is [13.62322 13.95355]. Meanwhile, the estimate with a stratification is 13.838 and 95% CI is [13.67224 14.00444]. By just comparing the results from two approaches, it is unclear if the stratification approach creates any adjustment.

(e) Examine the variances of `logsal` in each stratum using the population data. Do you think optimal allocation would be worthwhile for this problem? Why or why not?

Yes. It is worth taking the stratification approach to apply an optimal allocation to the sampling since each strata, which is the baseball team, has various means (=logsal) and variances. Stratification improves precision by creating subpopulations (=strata) to incorporate variations pertaining to stratum within their stratum respectively (=baseball team).