Stat410

Group Project

Estimate average grade of students in Beijing by using different sampling methods

Vicky Xu (301147374)

Lianzhuo Sun (301253184)

Introduction

Difference in grade can reflect difference in teaching quality. If the students are in the same grade level, it is often argued that the average in the same grade level should be the same. Many students and their parents believe that the schools with higher grade have higher teaching quality and they want to enroll in the schools with higher grade. As a result of that, in year 2015, the schools with higher grade were full of students and schools with lower grade enrolled in less students, which leads to a large unbalance in number of students enrolled in high school in Beijing.

In this project, we are going to use different sampling methods to try to estimate the average grade of students in Beijing, China. Since we are unable to get all the grade of all the students, we take the samples of the students in different schools in Beijing. And we use the sample data to estimate the average chemistry grade for the students. The dataset contain 41 classes in Beijing, China, average grade of 41 classes in and the number of students per class. The goal of the project is to use different sampling methods, to compare and decide which method of the mean is closest to the actual average grade for students in Beijing.

Method

In this project, we take both simple random sampling without replacement (SRS) and random sampling with replacement. Sampling with replacement means that taking n independent sample units form the population. For the simulation study, we set the parameters b= 10000 times in terms to collect the

simulation population. Based of SRS and PPS designs, we take sample n=10 from the population of N=41 for the following 5 different estimation methods to find the average grade for the students.

- Simple Random Sampling Without Replacement (SRS):

In this method, each sample values are not independent form each other. The samples are directly taking from the population.

- Unequal Probability Sampling:

There are two methods used under this design: Sampling with Replacement (PPS) and Hansen-Hurwitz Estimator (HH). The unequal probability design used in this problem with additional parameter Pi and πi, which is the number of students in each different school with different sizes. The sample units we taken form the population may be selected more than ones, because we collect the sample units under the replacement design.

- Ratio Estimation

In this method we add one more auxiliary variable r. The ratio estimation using the new variable r= average grade of students / average number of students. The auxiliary information is used both in the estimation and the design.

- Stratified Sampling

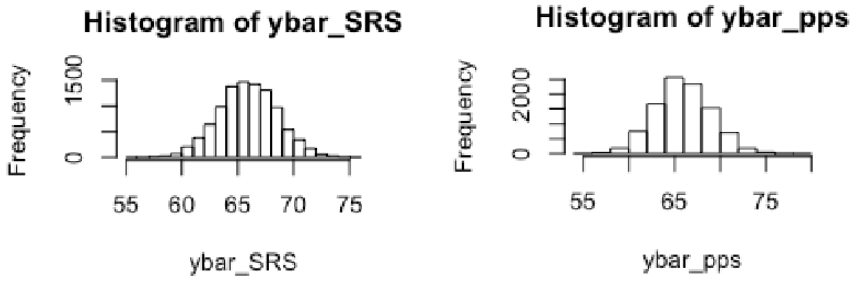
The rule of stratified sampling design is using SRS in each stratum in terms to keep the independence between each stratum. In this project problem, we choose three stratums by arranging the grades into three groups without the outliers. After excluding the maximum grade and the minimum grade, we obtain the three grade intervals are stratum1: [49.2562~60.31163], stratum2: [60.31163~71.0994], and stratum3: [71.0994~85.0423].

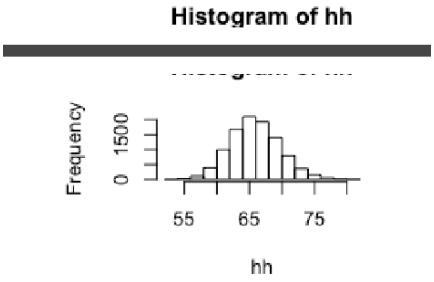
Result

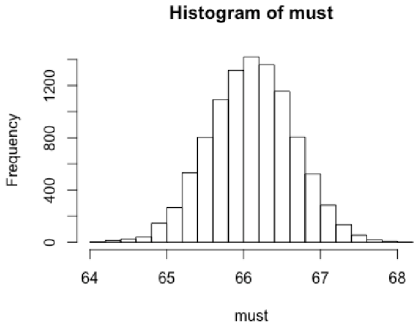
Result Table

Method/Result	Bias	MSE	Sample Mean
SRS	0.00024997	6.850928	65.90119
PPS	0.02635542	9.072034	65.9273
HH	-0.03435346	14.22243	65.86659
Stratified	-0.1181959	2.388734	65.78275
Ratio	0.101272517	10.417522	66.00222

According to the result table, the simple random sampling method has the smallest bias and the simple random sampling (SRS) method has the largest bias, but stratified sampling method has the smallest mean squared error value (MSE) and the Hansen-Hurwitz estimation (HH estimation) method producing the largest MSE value. The mean of average grade for all classes is 65.90094, the stratified sampling method has the largest value in comparison to the actual average grade 65.90094, The SRS estimation method produces the closest estimate to the actual average grade. Comparing the unequal probability sampling methods the HH estimator produces a highest MSE, the HH estimator is able to estimate more accurately the actual average grade of student for different classes, the HH estimator used for comparison with this unequal probability design.







According to the histograms for the generated simulation populations for each estimation method, all the histograms follow approximately a normal distribution with the histogram for the SRS, PPS method and HH estimator method being the closest to a symmetric bell curve. The SRS method has the smallest range for the estimated sample grade (55,75). The range for average grade in the histogram for the HH (55,80) explains the large MSE value.

Discussion:

From the result above, using simple random sampling method can estimate the most precise mean grade of student in Beijing.

However, the stratified sampling method shown that the high school classes in Beijing can be divided in three different levels. These three stratums are stratified by their mean grades, and

we are going to discuss the different quality of education that could be conducted from these classes between different stratums.

There is 12 classes in first stratum, which contains the classes with mean grades are between 49.26 and 60.31; the second stratum with mean grades between 60.31 and 70.10 contains 16 classes; the third stratum contains 12 classes with mean grade between 70.10 and 85.04.

We could conclude that the classes in first stratum have the lowest quality of education, also most of the school for these classes are located in rural area. The classes in third stratum with highest quality of education, their school are almost located in urban area. The school from urban area have the higher quality of education may because of the advanced development in economy, then they attach importance to education.