

Homework #2

Due October 15

1. An SRS of 290 households was taken from a city area with 14,828 households to estimate the proportion of households in that area that owned their home. Of the 290 households, 157 reported that they owned their home.

(a) Estimate the proportion of households in the city area that own their home and give a 99% confidence interval for this population proportion.

(b) Justify the assumptions necessary to obtain the confidence interval in part (a).

2. What is the minimum sample size necessary to estimate the proportion of people with blood type O in a population of 1800 people to be within 0.02 of the population proportion with 95% confidence? Assume no prior information about the proportion.

3. An unequal probability sample of size 3 is selected from a population of size 10 with replacement. The y -values and the selection probabilities are

y_i	p_i
5	0.08
12	0.025
9	0.16

(a) Estimate the population total using the Hanson-Hurwitz estimator.

(b) Estimate the variance of this estimator.

(c) Estimate the population total using the Horvitz-Thompson estimator.

(b) Give an unbiased estimate the variance of Horvitz-Thompson estimator.

You can use the Splus programs.

4. Consider the farm example. Assume that the number of farms and the area of the region is known. Estimate the proportion of farms with areas greater than 7 units by taking a sample of n farms with replacement using PPS. You need to define a variable y_i as follows:

$$y_i = \begin{cases} 1, & \text{if the area of farm } i \text{ is greater than 7} \\ 0, & \text{if the area of farm } i \text{ is not greater than 7.} \end{cases}$$

(a) What is the expected value of y_i (i.e., μ_y)?

(b) Use a sample of $n = 20$ farms to estimate the proportion of farms with areas greater than 7 units.

(c) Compute a 95% confidence interval for the true proportion.

You should use an Splus program; we will return to this problem soon.

5. From Barrett and Nutt, Survey Sampling in the Environmental Sciences, COMPRESS 1979, p. 211. Ecologists wish to estimate the number of species of breeding birds in a population of 38 recreational areas that had been constructed with as little disturbance of the natural

vegetation as possible. They took a random sample from the areas. The population units are 2-hectare plots and on each plot, they recorded the canopy diversity (x) and the number of breeding species (y). Canopy diversity is a measure of the coniferous and deciduous foliage about 12 feet, and is expressed as the percent of the less abundant component. The collected data are

Canopy Diversity	8	39	45	10	18	21	0	37	24	28	28	12	23	42	32
N. Species	9	13	10	8	9	10	4	10	11	12	8	8	8	14	11

The mean canopy diversity over the entire 38 areas was 26%. Estimate the mean number of breeding species using the sample mean and the ratio estimator. Discuss the how well the assumptions of the ratio estimators hold for these data. Calculate standard errors for the estimators and provide confidence intervals based on each estimate. Which method appears to be preferable?