SURVMETH / SURV 625
Applied Sampling
Winter / Spring 2025

Homework 3

1. The following are cluster totals y_α from $a = 20$ clusters of exactly $b = 10$ women (between the ages of 15 and 24) each. These clusters and the young women were sampled from the population frame used for Homework 1. The cluster totals y_α are the number of women who have ever been pregnant. Assume that the clusters were selected at random and *with replacement*, and the students were selected with *epsem* and *without replacement*. The sampling fraction is $f = ab/AB = n/N = 200/2{,}920 = 1 / 14.6$, meaning that <u>the finite population correction (fpc) should *not* be ignored in this case</u>.

α	1	2	3	4	5	6	7	8	9	10
y_α	4	4	3	6	4	6	3	4	4	1
α	11	12	13	14	15	16	17	18	19	20
y_α	1	8	3	3	5	6	4	5	8	5

(a) Compute an estimate of the mean $\bar{y}$, its standard error, and a 95% confidence interval for the population mean. (*Hint:* the degrees of freedom used in computing this confidence interval should not be 199.)

(b) Estimate the standard error of the mean that you would expect if the sample consisted of $a = 40$ clusters of $b = 10$ each. (*Hint:* What about this design has not changed, and what quantity needed to answer this question could therefore be considered portable?)

(c) Note that the mean $\bar{y}$ is a proportion. Based on the sample of 20 clusters [and ignoring the ratio $n / (n - 1)$], compute the design effect *deff,* as well as *roh*. How would you interpret the design effect for a colleague in plain English?

(d) Now, using the computed value of *roh* from part (c), estimate the standard error that you would expect from a sample of $a = 40$ clusters of $b = 5$ women each.

1

2. The data set, cherry.csv, contains measurements of diameter (inches), height (feet), and timber volume (cubic feet) for a sample of 31 black cherry trees. Diameter and height of trees are easily measured, but volume is more difficult to measure.

(a) Plot volume vs. diameter for the 31 trees.

(b) Suppose that these trees are an SRS from a forest of $N = 2967$ trees and that the sum of the diameters for all trees in the forest is 41,835 inches. Use ratio estimation to estimate the total volume for all trees in the forest. Give a 95% CI.

(c) Use regression estimation to estimate the total volume for all trees in the forest. Give a 95% CI.