Exercise ark #7.
Variance estimation

Oğuz–Alper, Melike & Pekarskaya, Tatsiana, Statistics Norway

October 9, 2020

Exercise 1

(R code available) Use the data in srs30.dat, which includes $n = 30$ units selected with simple random sampling from an artificial population of size $N = 100$.

1. Use the jackknife to estimate $V(\bar{y})$, and verify that $\hat{V}_{JK}(\bar{y}) = s^2/30$ for this data. What are the jackknife weights for jackknife replicate j?

2. Find also the bootstrap estimate of $V(\bar{y})$. (Lohr, 2019, p.394)

Exercise 2

(R code available) The file agsrs.dat contains data from an SRS of 300 of the 3 078 counties. Let y_i be total acreage of farms in county i in 1992 and x_i be total acreage of farms in county i in 1987. Use the jackknife and the bootstrap to estimate the variance of the ratio estimator $\hat{B}_r = \bar{y}/\bar{x}$. How do they compare with the linearization estimator? (Lohr, 2019, pp.394)

Exercise 3

(R code available) Foresters want to estimate the average age of trees in a stand. Determining age is cumbersome, because one needs to count the tree rings on a core taken from the tree. In general, though, the older the tree, the larger the diameter, and diameter is easy to measure. The foresters measure the diameter of all 1 132 trees and find that the population mean equals 10.3. They then randomly select 20 trees for age measurement.

Tree No.	Diameter, x	Age, y	Tree No.	Diameter, x	Age, y
1	12.0	125	11	5.7	61
2	11.4	119	12	8.0	80
3	7.9	83	13	10.3	114
4	9.0	85	14	12.0	147
5	10.5	99	15	9.2	122
6	7.9	117	16	8.5	106
7	7.3	69	17	7.0	82
8	10.2	133	18	10.7	88
9	11.7	154	19	9.3	97
10	11.3	168	20	8.2	99

Using the jackknife and the bootstrap, estimate the standard error for the regression estimate of the population age of trees in a stand. How do the jackknife and bootstrap compare with the standard error calculated using linearization methods? (Lohr, 2019, p.394)

Exercise 4

(R code available) The American Statistical Association (ASA) studied whether it should offer a certification designation for its members, so that statisticians meeting the qualifications could be designated as "Certified Statisticians." In 1994, the ASA surveyed its membership about this issue, with data in file certify.dat. The survey was sent to all 18 609 members; 5 001 responses were obtained. Results from the survey were reported in the October 1994 issue of Amstat News.

Assume that in 1994, the ASA membership had the following characteristics: 55% have PhD's and 38% have Master's degrees; 29% work in industry, 34% work in academia, and 11% work in government. The cross-classification between education and workplace was unavailable. There is a nonresponse. However, treat as if the respondents were selected with a probability sampling from the list of all members. Find the raking estimate of the total number of ASA members for this exercise. Estimate the total number of members opposing certification in 1994 using and use the bootstrap to estimate its variance and construct a 95% confidence interval. (Lohr, 2019, p.394)