STAT3003 Problem Sheet 4

ZHENG Weijia (William, 1155124322)

April 15, 2021

1 Q1

(Adapted from Exercise 11.10 from Scheaffer et al. (2012)) A corporation executive wants to estimate the proportion of corporation employees who have been convicted of a misdemeanor. Because the employees would not want to answer the question directly, the executive uses a random-response technique. A SRS of 300 people is selected from a large number of corporation employees. In separate interviews, each employee draw a card from a deck that has 0.7 of the card marked "convicted" and 0.3 marked "not convicted." The employee obtains 115 "yes" responses. Estimated the proportion of employees who have been convicted of a misdemeanor and provide a 95% confidence interval.

Solution

Let $\theta = 0.7$. And what we want to estimate is p.
Note that $P(Y) = 115/300 = \theta P(Y|C) + (1-\theta)P(Y|NC) 0.7P(Y|C) + 0.3P(Y|NC)$.
And note that $P(Y|C) = P(N|NC) = 1 - P(Y|NC)$.
Therefore $115/300 = 0.7P(Y|C) + 0.3(1 - P(Y|C))$. We solve out $P(Y|C) = 5/24$.
Hence $\hat{p} = P(Y|C) = 5/24 = 0.2083$.
And its variance is

$$\widehat{Var(\hat{p})} = \frac{1}{(2\theta-1)^2}(\frac{1}{n})P(Y)(1-P(Y)) = 4.9248 \times 10^{-3}.$$

Therefore a proper 95% C.I. for it can be

$$(\hat{p} \pm t_{n-1,1-\frac{\alpha}{2}}\sqrt{\widehat{Var(\hat{p})}}\,) = (0.2083 \pm 1.9679 \times 0.0702) = (0.2083 \pm 0.1381).$$

2 Q2

(Adapted from Exercise 5.21 from Scheaffer et al. (2012)) A quality control inspector must estimate the proportion of defective microcomputer chips coming from two different assembly operations. She knows that, among the chips in the lot to be inspected, 60% are from assembly operation A and 40% are from assembly operation B. In a SRS of 100 chips, 36 turn out be from operation A and 64 from operation B. Among the sampled chips from operation A, six are defective. Among the sampled chips from operation B, ten are defective.

(a) Considering only the SRS of 100 chips, estimate the proportion of defectives in the lot, and provide a 95% confidence interval.

(b) Stratifying the sample, after selection, into chips from operation A and B, estimate the proportion of defectives in the population, and estimate the standard error. Ignore the finite population correction in both cases. Which answers do you find more acceptable?

Solution

2.1 (a)

The point estimate should be $\hat{p} = \frac{6+10}{64+36} = \frac{16}{100} = 0.16$.
And using $\widehat{Var}(\hat{p}) = (1 - \frac{n}{N})\frac{1}{n-1}\hat{p}(1-\hat{p}) = 1.3575 \times 10^{-3}$.
Hence a proper 95% C.I. for it should be

$$(\hat{p} \pm t_{n-1,0.975}\sqrt{\widehat{Var}(\hat{p})}\ ) = (0.16 \pm 1.9842 * 0.0368) = (0.16 \pm 0.07301).$$

2.2 (b)

Use $\widehat{p_{pst}} = 0.6 * \frac{6}{36} + 0.4 * \frac{10}{64} = 0.1625$ as the point estimate.
We stratify all the chips in the lot into 2 groups: from A and from B. Therefore $\hat{p_1} = \frac{6}{36} = \frac{1}{6}$.
And $\hat{p_2} = \frac{10}{64} = 0.15625$.
Therefore, using $\hat{\sigma_i}^2 = \frac{n_i}{n_i-1}\hat{p_i}(1-\hat{p_i})$, one can have $\hat{\sigma_1}^2 = \frac{1}{7}$. And $\hat{\sigma_2}^2 = 0.1339$.
Using the formulae

$$\widehat{Var(\widehat{p_{pst}})} = \sum_{i=1}^{2} \frac{N-n}{nN}\frac{N_i}{N}\hat{\sigma_i}^2 + \sum_{i=1}^{2} \frac{1}{n^2}\frac{N-n}{N-1}(1-\frac{N_i}{N})\hat{\sigma_i}^2 = 1.4065 * 10^{-3}.$$

Then the s.e. is $\sqrt{1.4065 * 10^{-3}} = 0.03750$.
And a 95% C.I. for it should be

$$(0.1625 \pm 1.96 * 0.0375) = (0.1625 \pm 0.0735).$$

I think the latter answer is more acceptable because I think the point estimate if more reasonable, while these two's variance are almost the same.

3 Q3

(Adapted from Exercises 3.18 and 3.19 in Scheaffer et al. (2012)) Table 1 provides data for the 2001 school year on some K-12 education variables as well as populations for the New England states. For samples of $n = 2$ taken with probabilities proportional to the populations of the states, find all possible estimates of the total number of teachers in the New England states and demonstrate that the estimator is unbiased. Do this for

(a) Sampling with replacement.

(b) Sampling without replacement.

(c) Would the estimates change much if the sampling were done with probabilities proportional to the total number of students rather than to the state populations? Give a valid reason for your answer without calculating the actual estimates.

State	Total Students (1000)	Total Teachers (1000)	Expenditure per pupil	Population (100,000)
Connecticut	570	42	10,127	35
Maine	206	17	8232	13
Massachusetts	973	69	9509	64
New Hampshire	207	15	7286	13
Rhode Island	158	11	9315	11
Vermont	101	8	9153	6
Totals	2215	162		142

Table 1: State Education Data

Solution