1.14 Target population: All professors of education

Sampling population: List of education professors

Sampling unit: One professor

Information about how the sample was selected was not given in the publication, but let's assume it was a random sample. Obviously, nonresponse is a huge problem with this survey. Of the 5324 professors selected to be in the sample, only 900 were interviewed. Professors who travel during summer could of course not be contacted; also, summer is the worst time of year to try to interview professors for a survey.

1.15 Target population: All adults

Sampling population: Friends and relatives of American Cancer Society volunteers

Sampling unit: One person

Here's what I wrote about the survey elsewhere:

"Although the sample contained Americans of diverse ages and backgrounds, and the sample may have provided valuable information for exploring factors associated with development of cancer, its validity for investigating the relationship between amount of sleep and mortality is questionable. The questions about amount of sleep and insomnia were not the focus of the original study, and the survey was not designed to obtain accurate responses to those questions. The design did not allow researchers to assess whether the sample was representative of the target population of all Americans. Because of the shortcomings in the survey design, it is impossible to know whether the conclusions in Kripke et al. (2002) about sleep and mortality are valid or not." (pp. 97–98)

Lohr, S. (2008). "Coverage and sampling," chapter 6 of *International Handbook of Survey Methodology*, ed. E. deLeeuw, J. Hox, D. Dillman. New York: Erlbaum, 97–112.

2.1 (a) $\bar{y}_U = \dfrac{98 + 102 + 154 + 133 + 190 + 175}{6} = 142$

(b) For each plan, we first find the sampling distribution of $\bar{y}$.

Plan 1:

Sample number	$P(\mathcal{S})$	$\bar{y}_{\mathcal{S}}$
1	1/8	147.33
2	1/8	142.33
3	1/8	140.33
4	1/8	135.33
5	1/8	148.67
6	1/8	143.67
7	1/8	141.67
8	1/8	136.67

(i) $E[\bar{y}] = \frac{1}{8}(147.33) + \frac{1}{8}(142.33) + \cdots + \frac{1}{8}(136.67) = 142.$

(ii) $V[\bar{y}] = \frac{1}{8}(147.33 - 142)^2 + \frac{1}{8}(142.33 - 142)^2 + \cdots + \frac{1}{8}(136.67 - 142)^2 = 18.94.$

(iii) $\text{Bias}\,[\bar{y}] = E[\bar{y}] - \bar{y}_U = 142 - 142 = 0.$

(iv) Since $\text{Bias}\,[\bar{y}] = 0$, $\text{MSE}\,[\bar{y}] = V[\bar{y}] = 18.94$

Plan 2:

Sample number	$P(\mathcal{S})$	$\bar{y}_{\mathcal{S}}$
1	1/4	135.33
2	1/2	143.67
3	1/4	147.33

(i) $E[\bar{y}] = \frac{1}{4}(135.33) + \frac{1}{2}(143.67) + \frac{1}{4}(147.33) = 142.5.$

(ii)

$$
\begin{aligned}
V[\bar{y}] &= \frac{1}{4}(135.33 - 142.5)^2 + \frac{1}{2}(143.67 - 142.5)^2 + \frac{1}{4}(147.33 - 142.5)^2 \\
&= 12.84 + 0.68 + 5.84 \\
&= 19.36.
\end{aligned}
$$

(iii) $\text{Bias}\,[\bar{y}] = E[\bar{y}] - \bar{y}_U = 142.5 - 142 = 0.5.$

(iv) $\text{MSE}\,[\bar{y}] = V[\bar{y}] + (\text{Bias}\,[\bar{y}])^2 = 19.61.$

(c) Clearly, Plan 1 is better. It has smaller variance and is unbiased as well.

2.2 (a) Unit 1 appears in samples 1 and 3, so $\pi_1 = P(\mathcal{S}_1) + P(\mathcal{S}_3) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

Similarly,

$$
\begin{aligned}
\pi_2 &= \frac{1}{4} + \frac{3}{8} = \frac{5}{8} \\
\pi_3 &= \frac{1}{8} + \frac{1}{4} = \frac{3}{8} \\
\pi_4 &= \frac{1}{8} + \frac{3}{8} + \frac{1}{8} = \frac{5}{8} \\
\pi_5 &= \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\
\pi_6 &= \frac{1}{8} + \frac{1}{8} + \frac{3}{8} = \frac{5}{8} \\
\pi_7 &= \frac{1}{4} + \frac{1}{8} = \frac{3}{8} \\
\pi_8 &= \frac{1}{4} + \frac{1}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}.
\end{aligned}
$$

Note that $\sum_{i=1}^{8} \pi_i = 4 = n$.

(b)

Sample, $\mathcal{S}$	$P(\mathcal{S})$	$\hat{t}$
$\{1,3,5,6\}$	1/8	38
$\{2,3,7,8\}$	1/4	42
$\{1,4,6,8\}$	1/8	40
$\{2,4,6,8\}$	3/8	42
$\{4,5,7,8\}$	1/8	52

Thus the sampling distribution of $\hat{t}$ is:

k	$P(\hat{t} = k)$
38	1/8
40	1/8
42	5/8
52	1/8

2.4 (a)

The following, then, is the sampling distribution of $\bar{y}$.

k	$P(\bar{y} = k)$
$2\frac{1}{3}$	2/56
3	1/56
$3\frac{1}{3}$	4/56
$3\frac{2}{3}$	1/56
4	6/56
$4\frac{1}{3}$	8/56
$4\frac{2}{3}$	2/56
5	6/56
$5\frac{1}{3}$	7/56
$5\frac{2}{3}$	3/56
6	6/56
$6\frac{1}{3}$	6/56
7	1/56
$7\frac{1}{3}$	3/56

Using the sampling distribution,

$$E[\bar{y}] = \frac{2}{56}\left(2\frac{1}{3}\right) + \cdots + \frac{3}{56}\left(7\frac{1}{3}\right) = 5.$$

The variance of $\bar{y}$ for an SRS without replacement of size 3 is

$$V[\bar{y}] = \frac{2}{56}\left(2\frac{1}{3} - 5\right)^2 + \cdots + \frac{3}{56}\left(7\frac{1}{3} - 5\right)^2 = 1.429.$$

Of course, this variance could have been more easily calculated using the formula in (2.7):

$$V[\bar{y}] = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} = \left(1 - \frac{3}{8}\right)\frac{6.8571429}{3} = 1.429.$$

(b) A total of $8^3 = 512$ samples are possible when sampling with replacement. Fortunately, we need not list all of these to find the sampling distribution of $\bar{y}$. Let X_i be the value of the ith unit drawn. Since sampling is done with replacement, X_1, X_2, and X_3 are independent; X_i $(i = 1, 2, 3)$ has distribution

k	$P(X_i = k)$
1	1/8
2	1/8
4	2/8
7	3/8
8	1/8

Using the independence, then, we have the following probability distribution for $\bar{X}$, which serves as the sampling distribution of $\bar{y}$.

k	$P(\bar{y}=k)$	k	$P(\bar{y}=k)$
1	1/512	$4\frac{2}{3}$	12/512
$1\frac{1}{3}$	3/512	5	63/512
$1\frac{2}{3}$	3/512	$5\frac{1}{3}$	57/512
2	7/512	$5\frac{2}{3}$	21/512
$2\frac{1}{3}$	12/512	6	57/512
$2\frac{2}{3}$	6/512	$6\frac{1}{3}$	36/512
3	21/512	$6\frac{2}{3}$	6/512
$3\frac{1}{3}$	33/512	7	27/512
$3\frac{2}{3}$	15/512	$7\frac{1}{3}$	27/512
4	47/512	$7\frac{2}{3}$	9/512
$4\frac{1}{3}$	48/512	8	1/512

The with-replacement variance of $\bar{y}$ is

$$V_{\text{wr}}[\bar{y}] = \frac{1}{512}(1-5)^2 + \cdots + \frac{1}{512}(8-5)^2 = 2.$$

Or, using the formula with population variance (see Exercise 2.28),

$$V_{\text{wr}}[\bar{y}] = \frac{1}{n}\sum_{i=1}^{N}\frac{(y_i - \bar{y}_U)^2}{N} = \frac{6}{3} = 2.$$