MTH 432M: Introduction to Sampling Theory
Assignment No. 1: Sampling Designs

1. Consider a sampling design for sampling form a population $\mathcal{U} = (U_1, U_2, U_3)$ of three units, with study variables $\mathcal{Y} = (Y_1, Y_2, Y_3)$, and

$$P(\underline{s}) = \begin{cases} \frac{1}{7}, & \text{if } \underline{s} = (U_1, U_3) \\ \\ \frac{2}{7}, & \text{if } \underline{s} = (U_2, U_3) \\ \\ \frac{4}{7}, & \text{if } \underline{s} = (U_1, U_2, U_3) \end{cases} .$$

   Find

   (i) the first order inclusion probabilities π_1, π_2 and π_3.

   (ii) the second order inclusion probabilities π_{12}, π_{13} and π_{23}.

   (iii) the expected value and the variance of the estimator $\hat{\eta}(\underline{S}) = \frac{3}{n(\underline{S})} \sum_{i \in \underline{S}} Y_i$ of the population total.

2. Repeat Problem 1 with the sampling design described by

$$Q(\underline{s}) = \begin{cases} k, & \text{if } n(\underline{s}) = r(\underline{s}) = 2 \\ \\ 0, & \text{otherwise} \end{cases} ,$$

   where k is a fixed positive constant.

3. Let $A_i,\ i = 1, \ldots, N$, denote the number of times the i^{th} unit U_i appears in the sample drawn from the population $\mathcal{U} = (U_1, \ldots, U_N)$ under the $SRSWOR(n)$ design. Show that $\text{Cov}(A_i, A_j) = -\frac{n(N-n)}{N^2(N-1)},\ i \neq j$. Also find $\text{Cov}(A_i, A_j),\ i \neq j$, under the $SRSWR(n)$ design.

4. Let $\mathcal{Y} = ((X_1, Y_1), \ldots, (X_N, Y_N)),\ \overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i,\ \overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$ and $C_{\underline{X},\underline{Y}} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})$. Under the $SRSWR(n)$ design, let $\hat{\overline{X}} = \frac{1}{n} \sum_{i \in \underline{S}} X_i,\ \hat{\overline{Y}} = \frac{1}{n} \sum_{i \in \underline{S}} Y_i$ and $\hat{C}_{\underline{X},\underline{Y}} = \frac{1}{n-1} \sum_{i \in \underline{S}} (X_i - \hat{\overline{X}})(Y_i - \hat{\overline{Y}})$. Find

   (i) the correlation coefficient between $\hat{\overline{X}}$ and $\hat{\overline{Y}}$.

   (ii) $E(\hat{C}_{\underline{X},\underline{Y}})$.

5. For sampling from a population $\mathcal{U} = (U_1, \ldots, U_N)$ with the study variable $\mathcal{Y} = (Y_1, \ldots, Y_N)$, consider a sampling scheme under which $SRSWR$ is continued until the sample contains d (a fixed positive integer) distinct units. Let M denote the

number of selections made (i.e. M is the sample size, that is random) and, for $r = 1, \ldots, N$, K_r ($\sum_{r=1}^{N} K_r = M$) denote the frequency of the appearance of the r^{th} distinct unit in the sample. Define, $\hat{\overline{Y}}_1 = \frac{1}{M} \sum_{r \in \underline{S}} k_r Y_r$ and $\hat{\overline{Y}}_2 = \frac{1}{d} \sum_{r \in \underline{S}} Y_r$. Show that $V(\hat{\overline{Y}}_1) = \sigma^2 E(\frac{1}{M})$ and find $V(\hat{\overline{Y}}_2)$.

6. In a finite population $\mathcal{U} = (U_1, \ldots, U_N)$ of N units, let $\mathcal{Y} = (Y_1, \ldots, Y_N)$ be the study variable. Suppose that Y_1 is known. A $SRSWOR(n)$ is selected from the remaining units $(U_2, \ldots, U_N)$ and let $\hat{\overline{Y}}_{-1}$ be the sample mean of this sample. Let $\hat{\overline{Y}}$ denote the sample mean based on a $SRSWOR(n)$ from the entire population $\mathcal{U}$. Consider the following two estimators of the population total $T = \sum_{i=1}^{N} Y_i$.

   (i) $\hat{Y}_1 = Y_1 + (N-1)\hat{\overline{Y}}_{-1}$.

   (ii) $\hat{Y}_2 = N\hat{\overline{Y}}$.

   Are the above two estimators unbiased for estimating the population total T? Compare the variances of the above two estimators.

7. In a finite population $\mathcal{U} = (U_1, \ldots, U_N)$ of N units, let $\mathcal{Y} = (Y_1, \ldots, Y_N)$ be the study variable. Let Y_1 and Y_N be the extreme values such that the value of Y_1 is extremely low and the value of Y_N is extremely high among $Y_1, \ldots, Y_N$. Under the $SRSWOR(n)$ scheme, as an alternative to the sample mean $\hat{\overline{Y}}$, consider the following estimator for the population mean:

$$\tilde{\overline{Y}} = \begin{cases} \hat{\overline{Y}} + k, & \text{if the sample contains } U_1 \text{ but not } U_N \\ \hat{\overline{Y}} - k, & \text{if the sample contains } U_N \text{ but not } U_1 \\ \hat{\overline{Y}}, & \text{otherwise} \end{cases},$$

   where k is a fixed positive constant. Find $E(\tilde{\overline{Y}})$ and $V(\tilde{\overline{Y}})$.

8. Consider a $SRSWOR(2)$ from a population $\mathcal{U} = (U_1, U_2, U_3)$, with study variable $\mathcal{Y} = (Y_1, Y_2, Y_3)$. Consider the following estimator of the population mean:

$$\tilde{\overline{Y}} = \begin{cases} \frac{Y_1+Y_2}{2}, & \text{if the sample contains } U_1 \text{ and } U_2 \\ \frac{1}{2}Y_1 + \frac{2}{3}Y_3, & \text{if the sample contains } U_1 \text{ and } U_3 \\ \frac{Y_2+Y_3}{2}, & \text{if the sample contains } U_2 \text{ and } U_3 \end{cases}.$$

   Find $E(\tilde{\overline{Y}})$ and $V(\tilde{\overline{Y}})$.