**STAT 521: Homework Assignment 1**

**Due on Tuesday Jan. 30, 2018**

~~**Problem 1:** Identifying sample survey terms:~~

~~A survey is conducted to find the average weight of cows in a region. A list of all farms is available for the region, and 50 farms are selected at random. Then the weight of each cow at the 50 selected farms is recorded.~~

1. ~~What is the target population?~~
2. ~~What is the element?~~
3. ~~What is the sampling unit?~~
4. ~~What is the frame?~~
5. ~~List two possible sources of nonsampling errors.~~

~~**Problem 2:** For a fixed sample size design (i.e. $n$ is a fixed number), prove that the variance of the HT estimator can alternatively be written as~~

$$V(\widehat{Y}_{HT}) = -\frac{1}{2}\sum_{k\in U}\sum_{l\in U}\Delta_{kl}\left(\frac{y_k}{\pi_k}-\frac{y_l}{\pi_l}\right)^2,$$

WHAT THE HECK IS DELTA?

~~and its unbiased estimator is~~

Delta = pi_ij - pi_i*pi_j

$$\widetilde{V}(\widehat{Y}_{HT}) = -\frac{1}{2}\sum_{k\in S}\sum_{l\in S}\frac{\Delta_{kl}}{\pi_{kl}}\left(\frac{y_k}{\pi_k}-\frac{y_l}{\pi_l}\right)^2,$$

~~provided that all $\pi_{kl} > 0$.~~

**Problem 3**: Consider a population of size $N = 3$, $U = \{1, 2, 3\}$. Suppose $y_1 = 5$, $y_2 = 3$, and $y_3 = 7$. Consider the sampling scheme given in the table below.

| Sample | $A$ | $P(A)$ | HT est. of $t_y$ | HT Var. Est. |
|---|---|---|---|---|
| 1 | $\{1, 2\}$ | 0.4 | 10.53571 | |
| 2 | $\{1, 3\}$ | 0.3 | 17.91667 | |
| 3 | $\{2, 3\}$ | 0.2 | 15.95238 | |
| 4 | $\{1, 2, 3\}$ | 0.1 | 22.20238 | |

1. ~~Compute the first order inclusion probability, $\pi_k$, for each element, $k$.~~

2. ~~Compute the second order inclusion probabilities for each pair $k, \ell$ such that $k \neq \ell$.~~

3. ~~What is the expected sample size?~~

4. ~~What is the population total, $t_y$ for this population?~~

5. ~~Compute the HT estimator of $t_y$ for each of the four samples in the table above. Then, verify that $\hat{t}_{HT}$ is an unbiased estimator of $t_y$.~~

6. Compute the HT variance estimator for each of the four samples in the table above. Then, verify that $\hat{V}\{\hat{t}_{HT}\}$ is an unbiased estimator of $V\{\hat{t}_{HT}\}$.

7. Now consider estimating the proportion of the population with a value greater than or equal to 5. What is the population proportion? Using the sample in the first row of the table above, give the HT estimator of the proportion and a corresponding standard error.

8. Do you notice anything peculiar about the HT estimate for sample 4? Why/why not? For this design, can you define a different estimator of the total of $y$ that is unbiased for $t_y$ and has a smaller variance than the HT estimator.

**Problem 4:**

Consider a simple random sample without replacement $S$ of size $n$ from a finite population $U$ of size $N$, and two distinct individuals $k$ and $l$.

1. Compute $Pr\left(k \in S \text{ and } l \notin S\right)$

2. Suppose we select a further subsample $S_1$ from $S$ by simple random sampling of size $n_1$. Let $S_2$ be the complementary sample of $S_1$ in $S$. Thus, $S_1 \cup S_2 = S$ and $S_1 \cap S_2 = \phi$. Let $k$ and $l$ be any two distinct individuals belonging to the sample $S$. Compute $Pr\left(k \in S_1 \text{ and } l \in S_2\right)$.

3. Let $I_{k1} = I\left\{k \in S_1\right\}$ and $I_{k2} = I\left\{k \in S_2\right\}$, where $I\left\{A\right\}$ represents the indicator for event $A$. Compute $Cov\left\{I_{k1}, I_{l2}\right\}$.

4. Let $\bar{y}_i$ be the simple mean of $y$ in the sample $S_i$. Also, let $\bar{y}$ be the sample mean of $y$ in the sample $S$. Thus, $n\bar{y} = n_1\bar{y}_1 + n_2\bar{y}_2$ with $n_2 = n - n_1$. Prove that $Cov\left(\bar{y}_1, \bar{y}_2\right) = -S_y^2/N$ where $S_y^2 = \sum_{i=1}^{N}\left(y_i - \bar{Y}_N\right)^2/(N-1)$.

5. Under the above setup, compute $Cov\left(\bar{y}, \bar{y}_1 - \bar{y}\right)$ and the mean and variance of $\bar{y}_1$.