Instructions: Work any four problems. Show all work related to your solutions. You may use calculators and an 8.5 x 11 sheet with formulas.

1. An experienced farmer makes an “eye” estimate \( x_i \) of the weight of peaches on each tree in an orchard of \( N = 200 \) trees. He finds a total weight of \( t_{xU} = 11,600 \) pounds. The peaches are picked and weighed (with actual weights \( y_i \)) on a simple random sample of 10 trees with the following results.

\[
\sum_{i \in S} y_i = 543; \quad \sum_{i \in S} x_i = 569; \\
\sum_{i \in S} y_i^2 = 30483; \quad \sum_{i \in S} x_i^2 = 33227; \quad \sum_{i \in S} x_i y_i = 31794.
\]

The total weight of peaches was estimated by

\[
\hat{t}_y = t_{xU} + N(\bar{y}_S - \bar{x}_S).
\]

(a) Is this estimator unbiased? Estimate its standard error.

(b) Does it appear that a linear regression estimator using the sample least squares coefficient \( \hat{B}_1 \) would be more precise?

2. Two dentists, A and B, make a survey of the state of the teeth of 200 children in a village. Dr. A selects a SRS of 20 children and determines \( y \), the number of decayed teeth for each child, with the following results.

<table>
<thead>
<tr>
<th>No. of decayed teeth/child (( y ))</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of children</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Dr. B, using the same dental techniques, examines all 200 children, recording only those with no decayed teeth. He finds 60 children with no decayed teeth. Estimate the total number of decayed teeth among children in the village

(a) using only A’s results,
(b) using both A’s and B’s results.
(c) Are the estimates unbiased? Which do you expect to be more precise?
3. A simple random sample of households was chosen from a large population. The investigators recorded the number of persons in each sampled household and the number of members of the household who had visited a doctor at least once in the past year. Introduce suitable notation and propose an estimator $\hat{p}$ of the population proportion of persons who visited a doctor at least once in the past year. Also provide a variance estimate for $\hat{p}$.

4. A single cluster is drawn from a population of $N$ clusters with selection probabilities $\psi_i$, $i = 1, \ldots, N$.

(a) Prove that the statistic

$$\hat{y}_\psi = \sum_{i \in S} t_i/\psi_i$$

is an unbiased estimator of the population total $t_{y\ell}$, where $t_i$ is the $i$th psu total and all elements are observed in each selected psu.

(b) Prove that

$$\text{Var} (\hat{y}_\psi) = \sum_{i=1}^{N} \psi_i \left( \frac{t_i}{\psi_i} - t_{y\ell} \right)^2.$$ 

5. An investigator plans a stratified sample of a large population with two strata. It is known that sampling costs will be of the form $\sum c_h n_h$. His advance estimates of relevant quantities for the two strata are as follows.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>$W_h = N_h/N$</th>
<th>$S_h$</th>
<th>$c_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>10</td>
<td>$4$</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>20</td>
<td>$9$</td>
</tr>
</tbody>
</table>

(a) Find the values of $n_1/n$ and $n_2/n$ that minimize the sampling cost for a given value of $\text{Var} \hat{y}_{str}$, where $\hat{y}_{str}$ is the estimator of $\hat{y}_{\ell}$ under stratified sampling.

(b) Find the sample sizes required, under this optimal allocation, to make $\text{Var} \hat{y}_{str} = 1$. Ignore the finite population correction.

(c) Under this design, what is the variance of the estimated difference in stratum means?