# STAT 521: Midterm Exam Name:

**Problem 1:** (30 pts)

We are interested in estimating the proportion of international students at ISU with population size $N = 10,000$. Suppose that we have a simple random sample of size $n = 100$ and the result is as follows.

Table 1: Sample Distribution of Students

| | Male | Female |
|---|---|---|
| International | 30 | 20 |
| American | 20 | 30 |

It is known that the population proportion of male students is 60 %.

(a) Compute the 95% confidence interval for the proportion of international students at ISU using the post-stratified estimator with gender being the poststratum.

(b) Compute the reduction of variance due to poststratum compared to the HT estimator in the estimation of the proportion of international students at ISU.

(c) Suppose that we are interested in estimating $\theta = \theta_1 - \theta_2$, where $\theta_1$ is the proportion of international students among male students and $\theta_2$ is the proportion of international students among female students. Using the notation in the following table, $\theta_1 = N_{11}/N_{+1}$ and $\theta_2 = N_{12}/N_{+2}$. Using the sample in Table 1, estimate $\theta$ and its variance. (May ignore the finite population correction term.)

| | Sample | | Population | |
|---|---|---|---|---|
| | Male | Female | Male | Female |
| International | $n_{11}$ | $n_{12}$ | $N_{11}$ | $N_{12}$ |
| American | $n_{21}$ | $n_{22}$ | $N_{21}$ | $N_{22}$ |
| Total | $n_{+1}$ | $n_{+2}$ | $N_{+1}$ | $N_{+2}$ |

**Problem 2:** (20 pts)

Let $x_1, x_2, x_3, x_4, x_5$ be the five sample observations from SRS and we observed that $x_k = k$ in the sample. For this sample, we wish to assign the weights such that $\sum_{i=1}^{5} w_i = 1$ and $\sum_{i=1}^{5} w_i x_i = 4$. To uniquely determine $w_i$'s, suppose that we want to minimize

$$\sum_{i=1}^{n} \left( w_i - \frac{1}{n} \right)^2$$

subject to $\sum_{i=1}^{n} w_i = 1$ and $\sum_{i=1}^{n} w_i x_i = 4$, where $n = 5$. Find the resulting weights.

**Problem 3:** (30 pts)

Consider the following potential outcome model:

$$\begin{aligned} Y_i(1) &= \mathbf{x}_i'\boldsymbol{\beta}_1 + e_i(1) \\ Y_i(0) &= \mathbf{x}_i\boldsymbol{\beta}_0 + e_i(0) \end{aligned}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are unknown parameters, $e_i(1)$ and $e_i(0)$ are independent of $\mathbf{x}_i$ and $E\{e_i(1)\} = E\{e_i(0)\} = 0$. We obtain a realization of finite population of size $N$ from the above model and observe $(\mathbf{x}_i, T_i, Y_i)$ from the sample selected by completely randomized experiment, where

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0).$$

We are interested in estimating $\theta = E\{Y(1) - Y(0) \mid T = 1\}$, which is often called the average treatment effect on the treated (ATT). Note that the finite-population parameter for $\theta$ is

$$\bar{\theta}_N = \frac{1}{N_1} \sum_{i=1}^{N} T_i \left\{ Y_i(1) - Y_i(0) \right\}.$$

We are interested in using a weighted version of the estimator given by

$$\hat{\theta}_\omega = \sum_{i=1}^{N} T_i \omega_{1i} Y_i - \sum_{i=1}^{N} (1 - T_i) \omega_{0i} Y_i. \tag{1}$$

Answer the following questions.

(a) Find the conditions on the weights ($\omega_{1i}$ and $\omega_{0i}$) such that $\hat{\theta}_\omega$ in (1) is model-unbiased for $\bar{\theta}_N$.

(b) Find the optimal estimator that minimizes the model variance of $\hat{\theta}_{\omega}$ in (1) subject to the model-unbiasedness condition in (a). (May assume that $V\{e_i(1)\} = \sigma_1^2$ and $V\{e_i(0)\} = \sigma_0^2$. )

(c) Show that the optimal estimator in (b) is asymptotically design unbiased for $\bar{\tau}$, where

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \{Y_i(1) - Y_i(0)\} .$$

(May assume that $\mathbf{x}$ includes 1. )