MTH 432M: Introduction to Sampling Theory
Assignment No. 3: Stratified Random Sampling

1. Show that $\text{Var}_{Prop}(\hat{\overline{Y}}_{ST}) \leqslant \text{Var}_{Ran}(\hat{\overline{Y}})$ provided strata means are so different that
$$\sum_{h=1}^{L} N_h(\overline{Y}_h - \overline{Y})^2 > \sum_{h=1}^{L}(1 - \frac{n_h}{N_h})S_h^2.$$

2. Can the sample sizes under fixed sample size proportional allocation and fixed cost proportional allocation exceed corresponding strata size? Justify your answer either through counter examples or by providing proofs. What can you say about Neyman allocation and fixed cost optimum allocations?

3. If $2N$ population units are allocated to two strata of the same size ($N_1 = N_2 = N$). The total sample size $2n$ is allocated to strata in proportion to their sizes and SRSWOR is taken from each strata. Under what conditions $\text{Var}_{Prop}(\hat{\overline{Y}}_{ST}) \leqslant \text{Var}_{Ran}(\hat{\overline{Y}})$ holds?

4. A population is divided into two strata of sizes $2N_1$ and N_2. For the total sample size of $n = N_1 + N_2$, find conditions under which the optimum allocation is $n_1 = N_1$ and $n_2 = N_2$.

5. Suppose that study variable $Y \sim U(0, h)$, for some $h > 0$. The range $(0, h]$ is divided into L strata of equal sizes. A simple random sample of size $\frac{n}{L}$ is selected from each stratum. Let V_1 and V_2 denote the variances of the estimators of population mean for a simple random sample of size n and the stratified random sample, respectively. Show that $\frac{V_2}{V_1} = \frac{1}{L^2}$.

6. Consider a population $\mathcal{U} = (1, 2, \ldots, N)$ of size $N = Ln$ divided into L strata and h^{th} strata units labeled as $U_{hj} = (h-1)n + j,\ j = 1, \ldots, n, h = 1, \ldots, L$, where L and n are specified positive integers. From each stratum, only a random sample of size 1 is obtained. Suppose that the values of the study variable in the population are modeled as $Y_i = \alpha + \beta i, i = 1, \ldots, N$. Compare the variances of the unbiased estimators $\hat{Y}_{ST}$ and $\hat{Y}_{RAN}$ of the population total.

7. Consider a population of N units divided into two strata of sizes N_1 and N_2 ($N_1 + N_2 = N$). Suppose that the two strata have the same population variance. Consider a cost function $C = c_1 n_1 + c_2 n_2$, where c_1 and c_2 are the costs of sampling an observation from the first and the second strata respectively. Assuming that N_1 and N_2 are large, show that
$$\frac{V_{Prop}(\hat{\overline{Y}}_{ST})}{V_{Opt}\hat{\overline{Y}}_{ST}} = \frac{W_1 c_1 + W_2 c_2}{(W_1\sqrt{c_1} + W_2\sqrt{c_2})^2} \geqslant 1.$$

8. Consider a population of N units divided into two strata of sizes N_1 and N_2 ($N_1 + N_2 = N$). For a fixed sample size $2n$, let V_1 and V_2 denote the variances of $\hat{\bar{Y}}_{ST}$ under the equal allocation ($n_1 = n_2 = n$) and the optimum allocation. Let $\hat{n}_1$ and $\hat{n}_2$ denote the optimum sample sizes (under optimum allocation) and $\omega = \frac{\hat{n}_1}{\hat{n}_2}$. Assuming that N_1 and N_2 are large, show that

$$\frac{V_1 - V_2}{V_2} = (\frac{\omega - 1}{\omega + 1})^2.$$

9. Consider a population of N units divided into two strata of sizes N_1 and N_2 ($N_1 + N_2 = N$). Suppose that samples of sizes n_1 and n_2 are drawn from these strata. Let $\hat{n}_1$ and $\hat{n}_2$ denote the optimum sample sizes (under optimum allocation), $\tau = \frac{n_1}{n_2}$, $\omega = \frac{\hat{n}_1}{\hat{n}_2}$ and $\eta = \frac{\tau}{\omega}$. Let and V_1 and V_2 be the variances of stratum mean under actual and optimum allocations, respectively. Show that, when N_1 and N_2 are large,

$$\frac{V_2}{V_1} \geqslant \frac{4\eta}{(1+\eta)^2}.$$

10. Consider a population $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ with corresponding study variables $\mathcal{Y} = \{Y_1, Y_2, Y_3, Y_4\}$ given by $Y_1 = Y_2 = 0, Y_3 = 1$ and $Y_4 = -1$. Calculate the variance of the mean estimator for a $SRSWOR(2)$ design. Also, calculate the variance of the mean estimator for a stratified SRSWOR design with two strata $\mathcal{U}_1 = (U_1, U_2)$ and $\mathcal{U}_2 = (U_3, U_4)$, and with only one unit selected per stratum (proportional allocation). Analyse the findings.

11. Using the stratified sampling based on L strata, consider estimation of proportion of population of N units having a given attribute.

    (a) Derive the usual unbiased estimator $\hat{P}_{ST}$ of P, show unbiasedness, and derive its variance under SRSWOR within strata.

    (b) Ignoring fpc across strata, show that under proportional allocation ($n_h = nW_h$) the variance of $\hat{P}_{ST}$ reduces to that of SRS of size n from the whole population. Explain why proportional allocation is not always optimal.

    (c) Find Neyman allocation that minimizes variance of $\hat{P}_{ST}$ for fixed n. Give formula for n_h and optimum variance.

    (d) Given stratum costs c_h, overhead cost c_0 and total cost C, derive allocation minimizing variance under cost constraint $c_0 + \sum c_h n_h = C$. Under what conditions this allocation coincides with the Neyman allocation?

    (e) Derive a $100(1-\alpha)\%$ confidence interval for P using normal approximation.

    (f) Ignoring fpc across strata and assuming proportional allocation, find total sample size n so that $P(|\hat{P}_{ST} - P| \leqslant d) \geqslant 1 - \alpha$ (Hint: use CLT).

12. In stratified sampling from a population divided into two strata, the goal is to estimate the difference of population means of two strata. Find the optimum allocation of sample sizes n_1 and n_2 subject to the fixed cost $c = n_1c_1 + n_2c_2$.