Generalized Estimator on Operations Research Guided Sampling (ORGS)

***Pranjal Kaser**, Research Scholar, SoS in Statistics, Pt. RSU Raipur, C.G.
kaserpranjal@gmail.com

Abstract:

Survey sampling seeks to achieve statistical precision with limited resources. Classical approaches such as Neyman allocation minimize estimator variance under a fixed total sample size but often ignore heterogeneous costs, logistical constraints, and field feasibility. This paper introduces the **Operations Research Guided Sampling (ORGS)** framework, which integrates optimization techniques from Operations Research (OR) into traditional sampling theory to derive cost-efficient and operationally feasible sample allocations. By formulating sample allocation as a constrained optimization problem minimizing estimator variance subject to budget and feasibility limits- ORGS generalizes Neyman's allocation and yields the closed-form optimal solution $n_h^* = \frac{C, W_h S_h/\sqrt{c_h}}{\sum_{j=1}^{H} W_j S_j \sqrt{c_j}}$ where (c_h) denotes per-unit cost, (W_h) stratum weight, and (S_h) within-stratum standard deviation. The resulting minimum variance $V_{min} = \frac{\left(\sum_{h=1}^{H} W_h S_h \sqrt{c_h}\right)^2}{C}$ demonstrates how cost heterogeneity directly influences statistical efficiency.

Keywords: stratified sampling, optimal allocation, operations research, cost constraint, variance minimization, integer optimization.

1. Introduction

Survey sampling is one of the most powerful tools for extracting information about a population when a complete census is infeasible. The precision of a survey estimator often depends on how the total sample size n is allocated across sub-populations or strata. For a stratified random sample, the variance of the estimated mean can be expressed as

$$Var(\overline{y_{st}}) \approx \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{n_h}$$

where $W_h = N_h/N$ is the population weight of stratum h, S_h^2 is the within-stratum variance, and n_h is the sample size allocated to that stratum. The task of the statistician is thus to choose the $n_h's$ in a way that minimizes variance subject to certain restrictions.

Classically, Neyman (1934) showed that if the only restriction is a fixed total sample size $\sum_h n_h = n$, the optimal allocation is proportional to the stratum weight and variability:

$$n_h \propto W_h S_h$$

This allocation minimizes variance for a given sample size but ignores practical constraints such as differing costs of observation, geographical accessibility, or time limitations. In practice, surveys are rarely conducted in unconstrained environments. A sample unit from a remote tribal village, for example, may be far more costly than one from an urban ward. Likewise, certain areas may have limited accessibility due to terrain or political risk.

Operations Research (OR) provides a natural way to formalize such trade-offs. Instead of treating sampling as purely a statistical exercise, one can view it as an optimization problem under constraints. Suppose the per-unit cost of sampling in stratum h is c_h, and the total available budget is C. Then the sample allocation problem can be written as:

$$\text{Minimize V} = \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{n_h}$$

subject to the constraint

$$\sum_{h=1}^{H} c_h n_h \leq C, \quad n_h > 0$$

This formulation converts the sampling design into a constrained optimization problem. It can be solved using Operations Research tools such as Lagrange multipliers, linear programming, or stochastic optimization. Solving the above yields the Operations Research Guided Sampling (ORGS) allocation:

$$n_h^* = \frac{C\, W_h S_h / \sqrt{c_h}}{\sum_{j=1}^{H} W_j S_j \sqrt{c_j}}$$

with resulting minimum variance

$$Var_{min}(\overline{y_{st}}) = \frac{\left(\sum_{h=1}^{H} W_h S_h \sqrt{c_h} \backslash Big\right)^2}{C}$$

ORGS framework thus extends Neyman's allocation by incorporating real-world feasibility directly into the design. In this way, the strengths of sampling theory (statistical efficiency) and operations research (optimization under constraints) are combined into a unified paradigm.

This paper develops the theoretical basis of ORGS, demonstrates its proof, and provides a numerical illustration to highlight how it balances statistical precision with operational practicality. By introducing ORGS, we aim to offer a principled approach for modern survey designs where resource limitations cannot be ignored.

2. Review of Literature:

The problem of optimal allocation in stratified sampling has long been studied within both sampling theory and cost optimization frameworks. Classical works by Neyman (1934) and later extensions by Dalenius and Hodges (1959) focused on minimizing estimator variance under fixed sample size or cost constraints. These approaches, however, treat cost as a passive constraint rather than an active optimization parameter. Subsequent research introduced cost-based stratified allocation (Cochran, 1977; Sukhatme et al., 1984), where per-unit costs (c_h) are incorporated to adjust stratum sample sizes. The general form of this cost-sensitive allocation can be written as

$$\left[n_h \propto \frac{W_h S_h}{\sqrt{c_h}} \right]$$

which effectively balances cost and variability. Yet, this formulation assumes a single deterministic constraint and static conditions, lacking flexibility for multiple real-world restrictions such as time limits, accessibility, risk, and administrative feasibility.

Later developments in adaptive allocation and model-assisted designs (e.g., Rao, 2000; Beaumont, 2008; Lohr, 2010) allowed partial dynamic adjustment of sample sizes as auxiliary information evolved. However, these methods often operate sequentially and do not explicitly solve a global optimization problem—thus, they may improve estimation accuracy but cannot guarantee optimal resource utilization under competing constraints.

From the Operations Research perspective, several studies attempted to integrate optimization principles into sampling. Early attempts (Gunning & Horgan, 2004; Kozak, 2006; Khan et al., 2015) applied linear or nonlinear programming to minimize variance or cost. However, these studies typically framed the problem for specific cost functions or sampling goals, without

ISSN: 2583-6129
DOI:10.55041/ISJEM05915

establishing a unified, generalizable mathematical structure that links the optimization logic directly with the sampling variance function.

More recent work in adaptive and model-assisted designs (Rao, 2000; Beaumont, 2008; Lohr, 2010; Tille, 2020) focused on dynamic updating of sample sizes using auxiliary information. These methods improved estimation efficiency but often lacked a unified optimization framework that explicitly balanced statistical precision with operational feasibility. Similarly, stochastic and heuristic models (Khan et al., 2015; Singh et al., 2021) attempted to incorporate uncertainty in cost or variance but were often context-specific and computationally intensive.

3. Theoretical Framework (with equations):

It includes the Lagrangian derivation, the closed-form allocation (n_h^*) the resulting minimum variance, the special (Neyman) case:

Model and Solution

Consider the ORGS constrained optimization problem:

$$\text{Minimize:} \quad V = \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{n_h}$$

$$\text{subject to} \quad \sum_{h=1}^{H} c_h n_h \le C, \; n_h > 0 \quad h = 1, \dots, H,$$

where $(W_h = N_h/N)$, (S_h^2) is the within-stratum variance, (c_h) is the per-unit cost in stratum (h), and (C) is the total budget.

Lagrangian and first-order conditions

Form the Lagrangian (taking the budget as an equality at optimum):

$$\left[\mathcal{L}(n_1, \dots, n_H, \lambda) = \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{n_h} + \lambda! \left(\sum_{h=1}^{H} c_h n_h - C \right)\right]$$

where (λ) is the Lagrange multiplier for the budget constraint.
Differentiate $(\mathcal{L})$ with respect to (n_h) and set to zero:

$$\left[\frac{\partial \mathcal{L}}{\partial n_h} = -\frac{W_h^2 S_h^2}{n_h^2} + \lambda c_h = 0 \quad \Longrightarrow \quad n_h^2 = \frac{W_h^2 S_h^2}{\lambda c_h}\right]$$

Taking the positive root (since $(n_h > 0)$) gives

$$\left[n_h = \frac{W_h S_h}{\sqrt{\lambda}, \sqrt{c_h}}.\right]$$

Use the budget constraint $\left(\sum_{h=1}^{H} c_h n_h = C\right)$ to solve for $\left(\sqrt{\lambda}\right)$:

$$\left[\sum_{h=1}^{H} c_h \frac{W_h S_h}{\sqrt{\lambda}, \sqrt{c_h}} = C \quad \Longrightarrow \quad \frac{1}{\sqrt{\lambda}} \sum_{h=1}^{H} W_h S_h \sqrt{c_h} = C,\right]$$

hence

$$\left[\sqrt{\lambda} = \frac{\sum_{h=1}^{H} W_h S_h \sqrt{c_h}}{C}\right]$$

Optimal allocation

Substituting $(\sqrt{\lambda})$ back yields the closed-form optimal allocation:

This is the ORGS allocation: sample sizes are proportional to $(W_h S_h/\sqrt{c_h})$, scaled so that the cost constraint is satisfied.

Minimum achievable variance

Evaluating the objective at (n_h^*) gives the minimum variance:

$$\left[V_{min} = \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{n_h^*} = \sum_{h=1}^{H} \frac{W_h^2 S_h^2}{\frac{C W_h S_h/\sqrt{c_h}}{\sum_j W_j S_j \sqrt{c_j}}} = \frac{\left(\sum_{h=1}^{H} W_h S_h \sqrt{c_h}\right)^2}{C}\right]$$

So

$$\left[\boxed{V_{min} = \frac{\left(\sum_{h=1}^{H} W_h S_h \sqrt{c_h}\right)^2}{C}}\right]$$

special case - uniform costs (Neyman allocation)

If all costs are equal, $(c_h = c)$ for every (h), then $(\sqrt{c_h})$ cancels from the allocation and we obtain

$$[n_h^* \propto W_h S_h,]$$

which reduces to the classical **Neyman allocation** (up to the constant scaling that enforces $(\sum_h n_h = n)$ or $(\sum_h c_h n_h = C)$ Thus ORGS generalizes Neyman's rule by explicitly incorporating heterogeneous costs.

Results finding:

- The numerator $(W_h S_h)$ prioritizes strata with large size and high internal variability (as in Neyman).
- The division by $(\sqrt{c_h})$ down-weights expensive strata: even if a stratum is variable, its sample size is reduced when per-unit cost is high.
- (V_{min})ales inversely with total budget (C): doubling (C) (all else equal) halves the minimum variance.
- In practice, (n_h^*) may need rounding to integers and may be subject to additional feasibility bounds (minimum or maximum (n_h) per stratum). If such bounds exist, the problem becomes a constrained integer program.

4. Numerical illustration

Stratum (h)	**(W_h)**	**(S_h)**	**(c_h)**	**(n_h^*)**	**Rounded (n_h)**
1	0.50	10	100	38.37	38
2	0.30	15	225	23.02	23
3	0.20	8	64	15.35	15

Assume a population stratified into (H=3) strata. Take the following (made-up) values:

[$W = (W_1, W_2, W_3) = (0.50, ; 0.30, ; 0.20)$] (stratum weights, sum = 1)
[$S = (S_1, S_2, S_3) = (10, ; 15, ; 8)$] (within-stratum standard deviations)
[$c = (c_1, c_2, c_3) = (100, ; 225, ; 64)$] (per-unit costs)
[$C = 10{,}000$] (total budget)

Recall the ORGS formula:

$$n_h^* = \frac{C\, W_h S_h / \sqrt{c_h}}{\sum_{j=1}^{H} W_j S_j \sqrt{c_j}} \quad \text{And} \quad V_{min} = \frac{\left(\sum_{h=1}^{H} W_h S_h \sqrt{c_h}\right)^2}{C}$$

So the real-valued optimal allocation is approximately

$$(n_1^*, n_2^*, n_3^*) \approx (38.37, ; 23.02, ; 15.35),$$

with total (real) sample size ($\sum_h n_h^* \approx 76.746$).

minimum variance (V_{min})

on the formula:

$$V_{min} = \frac{(130.3)^2}{C} = \frac{16{,}978.09}{10{,}000} = 1.697809$$

So, the minimum achievable variance (under this model and budget) is ($V_{min} \approx 1.6978$)

5. References:

1. Alkaabneh, F., & Diabat, A. (2023). A multi-objective home healthcare delivery model and its solution using a branch-and-price algorithm and two-stage metaheuristic algorithm. Transportation Research Part C, 147, 103838.

2. Beaumont, J.F. (2008). *Model-assisted survey estimation with nonresponse and calibration*.

3. Bethel, J. (1989). *Sample Allocation in Multivariate Surveys*. Survey Methodology, 15, 47–57.

4. Chen, Y., & Wang, Y. (2023). Artificial intelligence and operations research in healthcare network operations: A review. Health Policy and Technology, 12(1), 125–134.
5. Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.
6. Dalenius, T., & Hodges, J. L., Jr. (1959). Minimum variance stratification. Journal of the American Statistical Association, 54(285), 88–101.

7. Gunning, P., & Horgan, J. M. (2004). Stratification methods. In Proceedings of the Federal Committee on Statistical Methodology (FCSM).
8. Khan, M. G. M. K., et al. (2015). Determining optimum strata boundaries and optimum allocation in stratified sampling. Aligarh Journal of Statistics, 35.

9. Kozak, M. (2006). "Optimal allocation in stratified sampling via nonlinear programming."

10. Lavallee, P. (2007) *Indirect Sampling.*

11. Lohr, S. L. (2010). *Sampling: Design and Analysis* (2nd ed.). Brooks/Cole..

12. Mahfouz, M. I., Rashwan, M. M., Khadr, Z. A., & Ramadan, M. A. (2024). "Multi-objective mathematical programming approach for multivariate compromise allocation for stratified random sampling.

13. Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97(4), 558–625.

14. Rao, J. N. K. (2000). *Small Area Estimation*. Wiley.

15. Rao, P. S. R. S. (2000). Sampling methodologies with applications. Chapman & Hall/CRC.
16. Restrepo, M. I., Rousseau, L. M., & Vallée, J. (2020). Home healthcare integrated staffing and scheduling. Omega, 95, 102057.
17. Sarndal, Swensson, & Wretman (1992) *Model Assisted Survey Sampling.*

18. Shad Muhammad, Y., Hussain, I., & Shoukry, A. M. (2016). "Multivariate Multi-Objective Allocation in Stratified Random Sampling: A Game Theoretic Approach." PMC.

19. Singh, et al. (2021). Spatial variability-based allocation. Computers & Geosciences.
20. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., & Asok, C. (1984). Sampling theory of surveys with applications (3rd ed.). Iowa State University Press.

21. Tille, Y. (2020). *Sampling Algorithms*. Springer.

22. Tillé, Y. (2020). Sampling and estimation from finite populations. Wiley.
23. Wang & Deng (2013) "*Multi-objective optimization for sample allocation.*" Applied Mathematical Modelling.
24. Yang, C. H., Chen, C. H., Hsu, W., & Chen, Y. H. (2023). Evaluation of smart long-term care information strategy portfolio decision model: The national healthcare environment in Taiwan. *Annals of Operations Research*, 1–32.