IJAER/July-August 2013 /Volume-02/Issue 4 ISSN: 2278-9677
International Journal of Arts & Education Research

BRIDGING THE GAP BETWEEN STATISTICAL CONCEPTS AND PRACTICAL IMPLEMENTATION

Sirajuddeen P K
Assistant Professor
Department Of Statistics
Govinda Pai Memorial Govt College, Manjeswar.
Kasaragod, Kerala

Abstract

Various tests are conducted for objectives such as quality assurance of items, buyer security, ecological monitoring, and interaction control, all of which require accurate and reliable data. The sampling theory of Pierre Gy provides a strong framework for enhancing scientific estimating norms and is regarded for its ability to manage diverse resources. Combined with stratified sampling, which divides the population into smaller groups for more accurate representation, this theory improves the precision and effectiveness of the entire logical interaction. Thorough improvement of the sampling and estimating phases ensures more accurate information assortment, reducing variability and improving overall outcomes dependability. Additionally, these enhanced methods can result in large expenditure investment money by reducing unnecessary sample efforts and optimising the estimate cycle. The integration of stratified sampling with Gy's theory provides a methodical approach to ensure that tests are astute and consistently produce excellent findings in a variety of application domains.

Keywords: *Statistical, Practical, Implementation, Sampling Theory, Stratified Sampling, Optimisation, Sampling, Bridging Gap.*

1. INTRODUCTION

The relationship between sampling methods and the reliability of scientific results is a cornerstone of statistical analysis and research integrity [1]. In many scientific disciplines, the emphasis on how representative samples is can be the difference between groundbreaking discoveries and misleading conclusions [2]. When researchers select samples that do not adequately represent the broader population or phenomenon being studied, they risk introducing biases that can skew results [3]. This concern highlights a persistent challenge within scientific literature: while there is ample discourse surrounding the importance of sampling, there remains a notable gap in actionable guidance for researchers [4]. The nuances of ensuring that a sampling system is robust often go unaddressed, leading to variability in how different studies achieve their findings [5]. Without a systematic approach to sample selection, researchers may inadvertently compromise the integrity of their findings, thus underscoring the necessity for clear, practical frameworks within the sampling methodology [6].

Sampling theory plays an essential role in bridging the conceptual gap between statistical principles and practical application in research [7]. This theoretical framework offers guidelines on how to achieve more reliable sample procedures, thus enhancing the validity of analyses across various scientific fields [8]. While the importance of sampling theory has long been recognized, its application has historically varied, leading to inconsistencies in research outcomes [9]. In recent years, however, there has been a noticeable shift among research institutions and professionals towards verifying and standardizing their sampling methods [10]. This trend is particularly evident in countries such as Finland, where there is an increasing awareness of the significance of employing scientifically sound sampling strategies [11]. The push towards standardization can be seen as a critical step in promoting best practices and fostering a more rigorous scientific environment, encouraging researchers globally to adopt these methods as they become more mainstream [12].

One of the pivotal concepts in achieving effective sampling is the standardization of methods, which can ultimately streamline the processes involved in data collection and analysis [13]. Standardizing sampling techniques ensures that researchers operate within a set of established parameters, leading to more uniform and replicable results across studies [14]. Key components of this standardization include the careful selection of sampling equipment, thorough evaluation of the methodologies' potential vulnerabilities, regular testing of sampling systems, and comprehensive training of personnel in executing these strategies [15]. By emphasizing these practical metrics, researchers can align their methodologies with theoretical principles of representativeness and statistical inconsistency.

2. LITERATURE REVIEW

M. F. Triola et al. (2004) In the book titled "Elementary Statistics," Triola and his colleagues give a foundational text that is intended to expose readers to the fundamentals of statistical analysis as well as its various applications. The fact that this work places a strong emphasis on real-world data and examples that encourage students and researchers to apply statistics to a wide variety of fields is one of the most prominent aspects of this work. Additional enhancements to the learning experience are provided by the incorporation of statistical software tools, which make it possible for readers to efficiently visualise and analyse data. The fact that this work continues to be widely used in educational institutions is evidence of the fact that it is successful in bridging the gap between theoretical comprehension and the practical application of statistics [16].

Wosten J. H. M. et al. (2001). The research that Wosten and associates conducted on Ped transfer functions (PTFs) addresses a significant barrier in the field of soil science. The estimation of soil hydraulic properties can be accomplished by the use of empirical or semi-empirical equations known as Ped transfer functions. As a result, predictions about irrigation efficiency, groundwater recharge, and soil erosion are improved as a result of this research. Particularly in places where field measurements are limited, such as developing countries like India, the authors' contribution to the development and improvement of PTFs has been significant in advancing the science of soil hydrology. This is especially true in regions where the scope of field observations is limited [17].

Allen et al. (2006) Within the context of official mentoring programs, Allen, Eby, and Lentz investigate the dynamics of mentorship. They concentrate on the quality of mentorship behaviours and the influence those behaviours have on both mentors and mentees simultaneously. On the other hand, mentors experience increased levels of job satisfaction and professional development. The purpose of this study is to highlight the significance of structured and intentional mentorship programs in order to bridge the gap between theoretical frameworks and real-world practices in a workforce that is rapidly globalising. This includes countries such as India, where mentorship is increasingly being integrated into corporate culture [18].

Nisbet et al. (2009). The "Handbook of Statistical Analysis and Data Mining Applications" written by Nisbet, Elder, and Miner is an exhaustive resource that is designed for scholars and practitioners that are active in data mining and analysis. It covers a wide variety of subjects, ranging from the fundamental principles of data mining to more sophisticated approaches such as machine learning and predictive analytics. The manual is especially helpful for individuals who are employed in fields that include the analysis of huge datasets, such as the healthcare industry, the marketing industry, and the financial sector. This handbook offers essential insights into the ways in which data mining may be utilised for decision-making, predictive analysis, and business intelligence. This is particularly important in light of the rapid progress of data-driven technologies in India's thriving information technology sector [19].

Claussen et al. (2002) published their findings. Within the context of the investigation of climate dynamics, Claussen et al. explore the development of Earth System Models of Intermediate Complexity (EMICs) as well as the significance of these models. EMICs offer a compromise between climate models that are extremely comprehensive and those that are conceptual because they provide a midway ground. This research comes at a time when countries like India are facing increasing challenges as a result of climate change. These challenges range from changed monsoon patterns to rising sea levels. Specifically, the study highlights the importance that EMICs play in giving a more detailed knowledge of the climate system of the Earth, which is essential for both scientific research and environmental policy [20].

3. PRACTICAL IMPLEMENTATION IN SAMPLING DESIGN AND AUDITING

In logical science, planning and evaluating sampling methods is essential to ensuring precise and reliable information comprehension. An organised technique that is rooted in both a theoretical comprehension of mistakes and practical contemplations is necessary to overcome any barriers that stand in the way of statistical theory's practical application. The sample errors list by Pierre Gy provides a thoughtful and clever framework for organising and analysing sampling procedures. This structure ensures the respectability of logical conclusions by identifying several sources of error and providing frameworks for limiting them in certified applications.

Step 1: Ensuring Correct Sampling Equipment and Procedures

Ensuring that all sample equipment and processes adhere to the criteria of proper sampling is the most crucial step in transitioning theory to practice. At this point, inappropriate sample tools and techniques should be identified and replaced with appropriate ones. This is consistent with Gy's theory, which emphasises that appropriate sampling may completely eliminate or reduce important errors like emergence and readiness problems. For instance, using inappropriate equipment might introduce bias and result in inaccurate representations of the content being dissected. A lab may prevent these tendencies by using the proper procedures, ensuring that the example is representative of the entire section.

Step 2: Estimating and Analysing Remaining Errors

Assessing the remaining errors, specifically the primary sampling error, the collecting and isolation error, and the point selection error, comes after the necessary sampling hardware and systems have been fixed. Since these errors are frequently more subtle, statistical techniques are needed to assess their impact. The magnitude of the additions and the frequency of sampling play a crucial role in the way these errors appear. The challenge is in adapting statistical models to real-world constraints in situations where perfect conditions might not always exist.

Step 3: Defining Acceptable Uncertainty Levels and Optimizing Procedure

The next phase is to describe the examination's acceptable degree of overall vulnerability or its financial constraints and, if necessary, adjust the sample plan. This entails determining the appropriate augmentation sizes and sample frequencies as well as applying statistical techniques to determine the optimal approach, such as exact or stratified sampling. Finding a balance between statistical rigour and practical believability is crucial in this situation. The goal is to achieve the required levels of vulnerability without causing needless costs or failures.

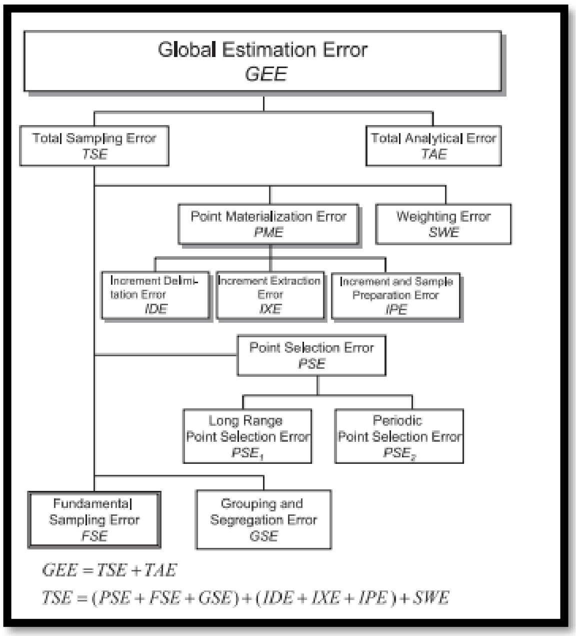

Figure 1: Hierarchical Representation of The Global Estimation Error (GEE) In Sampling and Analysis.

- **Preventing Bias and Ensuring Correctness**

The remaining pieces of Stage 1 are crucial in the meantime. When trends are provided, it might be difficult and costly to evaluate the risks of erroneous sampling. This emphasises the need for protection measures since inclinations are temporary and change with time, especially in situations where the material properties are different, such in stream confinement. On a fundamental level, statistical models explain how biases might distort discoveries, but in real life, it is undoubtedly more persuasive to prevent these biases with appropriate technology and techniques than to try to rectify them after the fact.

4. PRACTICAL IMPLEMENTATION IN FUNDAMENTAL SAMPLING ERROR MODELS

The Fundamental Sampling Error (FSE) highlights the basic error that can occur in an ideal sampling strategy and emphasises the real-world challenge of achieving precise results. The number of fundamental particles in the sample has a significant impact on the FSE, while this error is minimal for homogenous substances like gases and fluids. However, the FSE can be very important for solids, powders, and particulate materials, especially when the grouping of fundamental particles is small. This is similar to the difficulty of translating theoretical statistical models into real-world applications.

For example, statistical models like as the Poisson conveyance or binomial dissemination can be used to determine sampling susceptibility when the usual number of fundamental particles in an example can be easily determined. These models provide a statistical framework that enables experts to reduce intricate theoretical errors into meaningful information, providing a more useful and flexible method of handling dynamic in many contemporary cycles. Gy's main sampling error model is frequently still used as a foundational tool for evaluating change and effectively using sampling theory.

4.1.Estimation of Fundamental Sampling Error Using the Poisson Distribution

Poisson dispersion is a useful model for evaluating random distributions of rare events over an appropriate time or space interval, which makes it extremely relevant for real-world sampling. If the normal number of fundamental particles can be found, the Poisson dispersion may be used to deduce the example's standard deviation. Experts like Ingamells and Pitard have studied this strategy and demonstrated how to use it to evaluate sample errors. The indistinguishability of the mean and fluctuation of events within a certain length is a crucial characteristic of the Poisson dissemination.

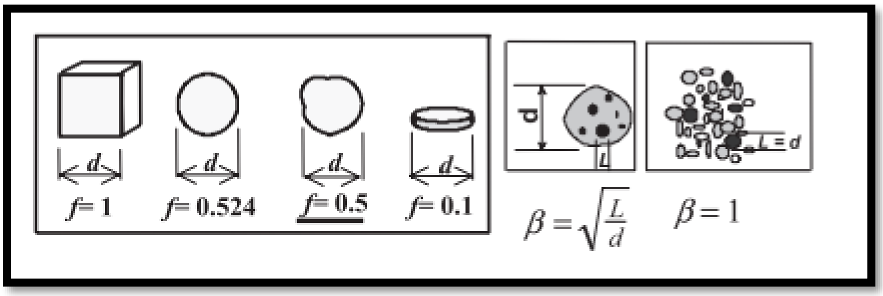

Figure 2: Particle Size Distribution and Shape Factors

For example: If one knows the typical number of fundamental particles (represented by λ), one can calculate the example's standard deviation as:

$$\sigma = \sqrt{\mu_n} \qquad (1)$$

Furthermore, the total standard deviation is basically

$$\sigma_r = \frac{1}{\sqrt{\mu_n}} \qquad (2)$$

When the usual number of particles (λ) is large (more than 25), the Poisson model may be approximated by typical circulation, which improves the certainty stretches. However, to accurately determine these stretches for smaller upsides of λ, Poisson dispersion itself should be used.

- **Example of Practical Application**

A plant director must ensure that about five particles larger than 5 μm are accessible per tonne of material in a fine-ground limestone creation office. The sample master assesses the predicted example size using the Poisson dispersion to ensure item quality. The master calculates that, given the supervisor's goal of achieving a 20% relative standard deviation, around 25 particles are needed in this scenario. Given that the package only has five coarse particles per tonne, it follows that five tonnes would be the critical example in order to get the required accuracy.

Relative standard deviation at its highest point is s_r = 20% = 0.2. We can determine how many coarse particles the example should include in order for this standard deviation to exist using Equation (2).

$$n = \frac{1}{s_r^2} = \frac{1}{0.2^2} = 25$$

Nevertheless, collecting and analysing a 5-ton test isn't feasible in practice, exposing a common disconnect between statistical theory and actual application. Instead of relying solely on traditional methods for sampling and inquiry, the master suggests an optional approach that revolves around process management. The plant may achieve the optimal item quality without the need for costly and inefficient sampling approaches by maintaining hardware execution and ensuring continuous creation.

5. OPTIMIZATION OF SAMPLING PLANS BASED ON STRATIFIED SAMPLING

Sommer and Cochran explored stratified sampling as a crucial strategy for improving sample designs. When there are standard layers present, such as in situations with packs, holders, or cart loads, this process is very effective. As demonstrated by Gy and Sommer, stratified sampling frequently results in reduced vulnerability concerning mean value, which can be on par with or superior to that achieved by random sample. This method increases appraisal accuracy and removes any obstacles preventing the use of theoretical statistical principles in real-world settings, enabling more potent dynamics in a variety of domains, such as quality control and exploration.

5.1. Optimization of Hierarchical Sampling Plans for Uniformly Sized Strata

Figure 3 illustrates the finalised sampling strategy, which calls for collecting samples at three successive levels (k = 3). The overall susceptibility of the parcel's mean increases with each step. The example from the upper level is used as the parcel for that level in the sampling chain at every subsequent level below the first.

5.1.1. Lot Characteristics

The component is composed by N_1 layers, or sublots, of corresponding sizes. n_1 layers are selected for sampling from these. The variation between the methods for the N_1 layers is measured by the standard deviation σ_2 . Choosing a layer often has a unit cost of c_2, which is negligible because it essentially includes direction.

5.1.2. Primary Samples

We extract n_2 key instances from each selected layer. The possible instances that can be obtained are addressed by the size of each layer n_2 .The standard deviation $\sigma_2$indicates the discrepancy between the crucial instances (within the layers standard deviation), whereas c_2 represents the cost associated with selecting a crucial example**.**

5.1.3. Analytical Samples

At this stage, each essential example has n_3 insightful examples available, with n_3 illuminating the potential scientific examples from an essential example. The unit cost for building up each scientific example is represented by c_3 and the standard deviation σ_3 reflects the variation in planning logical instances**.**

- **Cost and Variance Calculations**

Unit costs c_i can be expressed in terms of money or related expenses, such as time, for optimisation purposes. Due to autocorrelation in sample units and layers, which is a common in-process inspection, differences should ideally be evaluated using Gy's variographic approach. Even though fluctuation analysis is frequently used to evaluate different sections, it should only be used in situations where there is no autocorrelation or when sampling is strictly random.

- **Mean and Variance of the Lot**

An unweighted mean of the scientific results may be used to get the mean of the portion, and it is expressed as:

$$\text{Total number of samples analyzed : } n_t = n_1 n_2 n_3$$

$$\text{Mean of the lot : } \bar{x} = \frac{1}{n_1 n_2 n_3} \sum_i^{n_1} \sum_j^{n_2} \sum_k^{n_3} x_{ijk}$$

Variance of the lot mean:

$$\sigma_{\bar{x}}^2 = \frac{N_1 - n_1}{N_1 - 1} \frac{\sigma_1^2}{n_1} + \frac{N_2 - n_2}{N_2 - 1} \frac{\sigma_2^2}{n_1 n_2} + \frac{N_3 - n_3}{N_3 - 1} \frac{\sigma_3^2}{n_1 n_2 n_3}$$

Eq. when's what the difference indicates: the mean fluctuation completely eliminates the between-layers variation in the unlikely event when a sample can be selected from each layer ($N_1 = n_1$). However, if the example at all levels differs significantly from the parcel from which it is drawn, then this condition reverts to

$$\sigma_{\bar{x}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3} \quad \text{if all } n_i \ll N_i$$

$$\text{Total cost of the investigation : } c_t = n_1 c_1 + n_1 n_2 c_2 + n_1 n_2 n_3 c_3$$

- **Optimization Approaches**

There are two methods to improve this framework: either fix the all-out cost and restrict the mean fluctuation, or limit the total cost while figuring out a most extreme mediocre difference. Even though the need for numerical examples makes a precise mathematical arrangement difficult, optimisation can be achieved by involving inexact arrangements as demonstrated by Sommer or by conducting thorough checks of feasible arrangements, which are supported by current computational power. The imprecise arrangement acknowledges that $N_i \gg n_i$ is more significant.

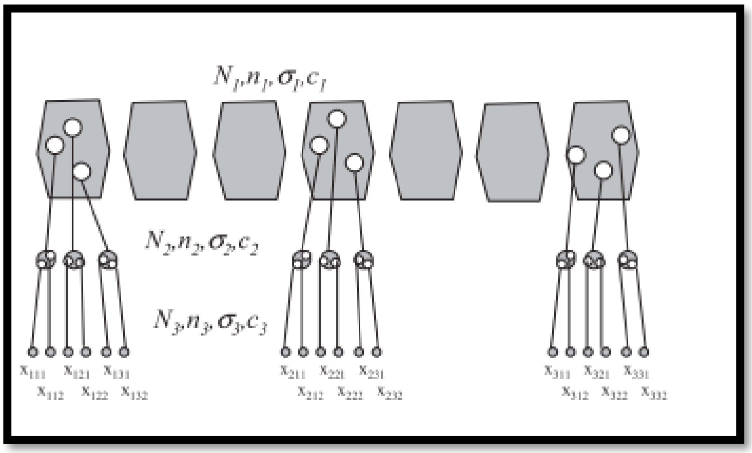

Figure 3: Hierarchical Representation of a Sampling Plan with Multiple Levels

5.1.4. Fixed Maximum Costs and Variance Minimization

The equation may be used to calculate the number of tests to be done at different levels when updating sampling programs under fixed highest expenditures (C_{max}):

$$n_i = \frac{s_i}{s_{i-1}} \sqrt{\frac{c_{i-1}}{c_i}}, \text{ constrained to integers } 1 \leq n_i \leq N_i,$$

for levels $i > 1$

By substituting the values for n_i (i>1) in Eq. (9), n_1 can now be solved

$$n_1 = \frac{c_{\max}}{c_1 + n_2 c_2 + n_2 n_3 c_3}, \text{ rounded to the lower intege}$$

$$1 \leq n_1 \leq N_1$$

5.1.4.1.Fixed Target Variance and Cost Minimization

When the mean's objective change, σ_T^2, is predetermined, the sampling convention should ensure that the methodology's fluctuation satisfies the requirement $\sigma_T^2 \leq \sigma_T^2$. For levels more than one, the previously stated equation is applicable.

$$n_1 = \frac{\dfrac{N_1}{N_1-1}\sigma_1^2 + \dfrac{N_2-n_2}{N_2-1}\dfrac{\sigma_2^2}{n_2} + \dfrac{N_3-n_3}{N_3-1}\dfrac{\sigma_3^2}{n_2 n_3}}{\sigma_T^2 + \dfrac{\sigma_1^2}{N_1-1}}$$

$$n_1 = \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{\sigma_T^2}$$

These solutions have to be rounded to the nearest upper integer $1 \leq n_1 \leq N_1$.

6. THE DETERMINATION OF COBALT CONTENT IN NICKEL CATHODES

Ensuring that nickel cathodes contain the appropriate amount of cobalt is an essential step, especially when certain resistances are imposed for cobalt levels, such as not exceeding 150 g/ton. In this model, we study an improved logical technique designed to ensure that the parcel mean's standard deviation does not exceed 5 g/ton. The scenario involves capturing instances at three distinct levels of error age, with the aim of minimising expenses while maintaining accuracy.

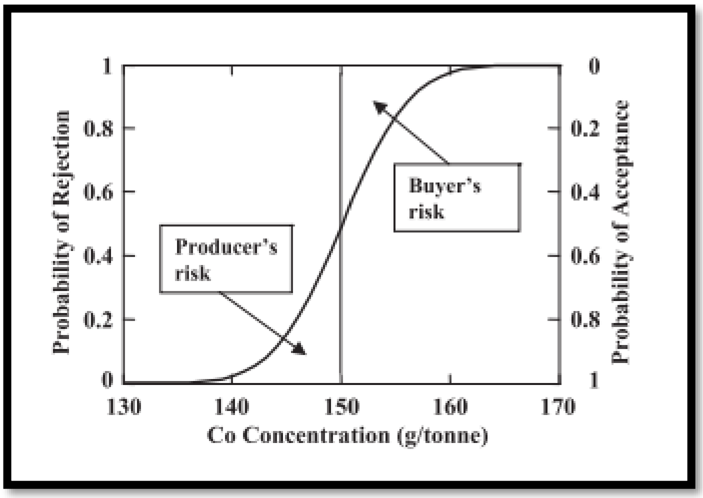

Figure 4: Operating Characteristic Curve (OC Curve)

- **Problem Statement**

The lot of cathode nickel weighs 25 tonnes, and the cobalt content should not exceed 150 g/tonne. The average weight of cathode plates produced is 50 kg, and these are further cut into 50-g pieces before packaging. These 50-g pieces are treated as the primary samples for cobalt determination, and a 1-g sample from each 50-g piece is dissolved for analysis. The cost for one analytical determination is 12 o, while taking one primary sample from a given plate cost.

The goal is to determine the number of samples required at each stage of the sampling process to ensure the standard deviation of the lot mean remains below 5 g/tonne.

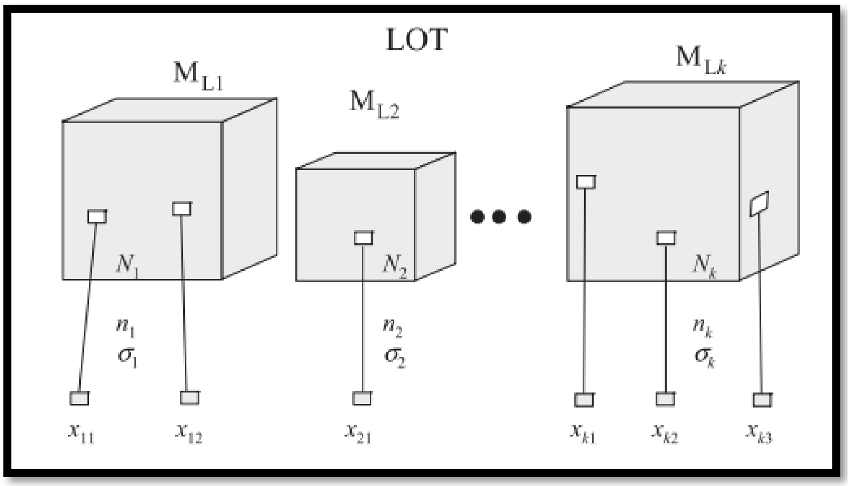

Figure 5: Multi-Stage Sampling Scheme

- **Error Generation Levels and Parameters**

Sampling is performed at three error-generating levels:

Table 1: Summary of Sampling Levels and Parameters

Level	**Description**	**Unit Size**	**N (Units)**	**Standard Deviation (g/tonne)**	**Cost (o)**	**Sample Size (n)**
Level 1: Between-Plate Variation	50-kg plate	50 kg	N1 = 500	s1 = 35	c1 = 0	(n1)
Level 2: Within-Plate Variation	50-g pieces cut from each plate	50 g	N2 = 1000	s2 = 15	c2 = 2	(n2)
Level 3: Analytical Determination	1-g sample dissolved from each 50-g piece	1 g	N3 = 50	s3 = 3.3	c3 = 12	(n3)

Optimization Strategy

The idea is to keep overall sample costs to a minimum while ensuring that the component mean's standard deviation falls below the acceptable upper bound of 5 g/ton. The number of tests taken at each stage and the associated costs affect the cycle's overall cost.

Using the provided equation for optimization, the following results were obtained:

- At **Level 3**, it was determined that the number of samples n_3 should be 1. This means that only one 1-g sample will be taken from each 50-g piece for cobalt analysis.
- At **Level 2**, similarly, the number of samples n_2 was set to 1, indicating that only one 50-g piece will be taken from each 50-kg plate for the analysis.
- At **Level 1**, the number of plates to be sampled n_1 was calculated as 53.3, or rounded to 55 plates, as per the equation applied.

Sampling Protocol

Given the calculations, the sampling protocol that could be followed is as such:

- Take a 50-g piece from every ninth cathode plate during the packing stage.
- This results in a total of 55 samples per 25-tonne lot.
- From each of these 50-g pieces, one cobalt determination is performed.

Cost and Precision

The total cost of this sampling protocol is calculated as:

$$55\times (0+2+12) =770$$

The expected standard deviation of the lot mean is approximately 4.9 g/tonne, which is below the target limit of 5 g/tonne, thus meeting the required precision for cobalt determination.

7. CONCLUSION

Sampling theory is fundamental to enhancing scientific approaches, yet it is frequently ignored in several fields, such as ecology and quality control. Typically, sample strategies are developed without careful evaluation, which leads to flaws and improperly opened opportunities for optimisation. Appropriate application of sampling theory may prevent unnecessary expenses, beginning with well-defined vulnerability targets. expenses are quadrupled by halving the standard deviation, and expenses are drastically increased by additional reductions; nevertheless, these investment funds may be justified if the logical cycle is strengthened. Effective resource allocation ensures that even the highest levels of vulnerability are satisfied without needless expense. In one scenario, a mash processing factory discovered that it was miscalculating the amount of mash delivered by up to 10% due to sampling and adjustment errors, resulting in significant financial losses. By improving the sample and alignment procedure, accurate errors were eliminated and random errors were restricted. This reduced the impact on the average annual mass of mash to just tenths of a percent, increasing accuracy and lowering costs.

REFERENCES

[1] Shields, P. M. (2006). Using pragmatism to bridge the gap between academe and practice.

[2] Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., ... & Saul, J. (2008). Bridging the gap between prevention research and practice: the interactive systems framework for dissemination and implementation. *American journal of community psychology*, *41*, 171-181.

[3] Breyfogle III, F. W. (2003). *Implementing six sigma: smarter solutions using statistical methods*. John Wiley & Sons.

[4] Holmström, J., Ketokivi, M., & Hameri, A. P. (2009). Bridging practice and theory: A design science approach. *Decision sciences*, *40*(1), 65-87.

[5] Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.*, *1*(1), 451-488.

[6] Fausch, K. D., Torgersen, C. E., Baxter, C. V., & Li, H. W. (2002). Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes: a continuous view of the river is needed to understand how processes interacting among scales set the context for stream fishes and their habitat. *BioScience*, *52*(6), 483-498.

[7] Coyne, J. C., & Racioppo, M. W. (2000). Never the Twain shall meet? Closing the gap between coping research and clinical intervention research. *American psychologist*, *55*(6), 655.

[8] Elmore, R. F. (2002). Bridging the gap between standards and achievement: The imperative for professional development in education. *Secondary lenses on learning participant book: Team leadership for mathematics in middle and high schools*, 313-344.

[9] Eagle, K. A., Garson Jr, A. J., Beller, G. A., & Sennett, C. (2003). Closing the gap between science and practice: the need for professional leadership. *Health Affairs*, *22*(2), 196-201.

[10] Starkey, K., & Madan, P. (2001). Bridging the relevance gap: Aligning stakeholders in the future of management research. *British Journal of management*, *12*, S3-S26.

[11] Dąbrowski, A. (2001). Adsorption—from theory to practice. *Advances in colloid and interface science*, *93*(1-3), 135-224.

[12] Jackson, A. W., & Andrews, G. A. (2000). *Turning points 2000: Educating adolescents in the 21st century*. Teachers College Press.

[13] Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications*. John Wiley & Sons.

[14] Forrester, A., Sobester, A., & Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.

[15] Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, *31*(3), 328-349.

[16] Triola, M. F., Goodman, W. M., Law, R., & Labute, G. (2004). *Elementary statistics* (p. 794). Boston: Pearson/Addison-Wesley.

[17] Wösten, J. H. M., Pachepsky, Y. A., & Rawls, W. J. (2001). Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of hydrology*, *251*(3-4), 123-150.

[18] Allen, T. D., Eby, L. T., & Lentz, E. (2006). Mentorship behaviors and mentorship quality associated with formal mentoring programs: closing the gap between research and practice. *Journal of applied psychology*, *91*(3), 567.

[19] Nisbet, R., Elder, J., & Miner, G. D. (2009). *Handbook of statistical analysis and data mining applications*. Academic press.

[20] Claussen, M., Mysak, L., Weaver, A., Crucifix, M., Fichefet, T., Loutre, M. F., ... & Wang, Z. (2002). Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate dynamics*, *18*, 579-586.

******