DOI: https://doi.org/10.57233/ijsgs.v10i2.651 ISSNp: 2488-9229; ISSNe: 3027-1118

IJSGS FUGUSAU VOL10 (2) WEBSITE: https://fugus-ijsgs.com.ng

INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY

(A PUBLICATION OF FACULTY OF SCIENCE, FEDERAL UNIVERSITY GUSAU, NIGERIA)

A Comparison between Regression and Ratio Estimators using Auxiliary Information: A Case Study of Ladoke Akintola University of Technology, Nigeria

[1]Lukman Abiola Oladimeji, [1]Samuel Abayomi Oke, [1]Oludayo Olugbenga Akinade, [1]Ibrahim Olalekan Ismail and [2]Olayemi Joshua Ibidoja

[1]Department of Statistics, Ladoke Akintola University of Technology Ogbomoso Nigeria

[2]Department of Mathematics, Federal University Gusau, Gusau, Nigeria

Corresponding Author's Email: ojibidoja@fugusau.edu.ng

Received on: April, 2024 **Revised and Accepted on:** May, 2024 **Published on:** July, 2024

ABSTRACT

This study investigated the use of separate and combined stratified random sampling to estimate population mean scores for two important courses in Statistics using their pre-requisites, by comparing two estimators (ratio and regression). For Probability Distribution, the separate ratio estimator emerged as the optimal choice, providing a mean score estimate of 31.66 with a variance of 6612.18. This indicates that using the pre-requisites course score as auxiliary information improved the estimation accuracy compared to the combined ratio estimator. In contrast, Statistical Inference, the combined ratio estimator proved to be more effective, yielding a mean score estimate of 33.99 with a variance of 84.54. The separate regression estimator was also evaluated, demonstrating its suitability for Probability Distribution with a mean score estimate of 32.18 and a variance of 7.22. However, for Statistical Inference, the combined regression estimator offered a more precise estimate (32.99) with a lower variance (17.59). The findings highlight the effectiveness of auxiliary information when selecting estimation methods in stratified sampling.

Keywords: Ratio, Regression, Pre-requisites, Auxiliary Information, Optimal, Stratified Sampling

1.0 Introduction

Educational systems are structured in a procedural manner, with lower-level courses serving as the foundation for higher-level courses (Alyasin et al., 2023). Pre-requisite courses represent the lower-level courses that provide baseline knowledge and skills required for follow-on core courses (Islam et al., 2008; Rach & Ufer, 2020; Shaffer et al., 2018). Statistically, pre-requisite courses can be viewed as independent variables that influence and help predict outcomes in core courses, which are the dependent variables. The performance and grades in a pre-requisite course provide auxiliary data that can inform estimates of central tendency (e.g. mean scores) and variation (e.g. variance) for student achievement in core courses (Ahmad, Ullah, et al., 2023; Ahmad, Zahid, Shabbir, Aamir, & Onyango, 2023; Ahmad, Zahid, Shabbir, Aamir, Emam, et al., 2023; Aslam et al., 2023; Balogun et al., 2020; Ibidoja et al., 2016; Kumar & Choudhary, 2023). Pre-requisite course grades demonstrate qualification for enrolment in core courses (Hands & Limniou, 2023).

In survey sampling, estimating population parameters often relies on information gathered from a limited sample (Bee et al., 2010; López, 2023; Pal & Singh, 2022; Srivastava, 1967). Ratio and regression estimators utilize auxiliary information, known for some or all population units, to improve the accuracy and efficiency of these estimates (Subzar et al., 2023). This review explores both separate and combined approaches, examining their strengths, weaknesses, and comparisons in the literature. The ratio estimator leverages a strong correlation between the study variable (y) and an auxiliary variable (x) by calculating the ratio of their population means ([R = Y/X]). It then applies this ratio to the sample mean of the auxiliary variable ($\bar{x}$) to estimate the population mean of the study variable ($\hat{y} = R * \bar{x}$). This method performs well when the ratio is constant across the population (Särndal et al., 1992a). Regression Estimator approach utilizes a linear regression model to estimate the study variable based on the auxiliary variable ($\hat{y} = \beta_0 + \beta_1 * x$). By leveraging the relationship between y and x, it can offer improvements compared to the simple mean estimator, especially when the relationship is non-linear (Lohr, 2010; Theobald, 2014).

Combined ratio estimator with a regression adjustment term ($\hat{y} = R * \bar{x} + b(\bar{x} - X)$). This can be beneficial when the ratio varies across subpopulations but the relationship between y and x within subpopulations remains linear (Nair, 1988). Regression-Ratio Estimator Swaps the order of operations, first regressing y on x to obtain residuals (e), then calculating the ratio of their means ([R' = E/X)] and applying it to the sample mean of the auxiliary variable ($\hat{y} = R' * \bar{x}$). This can be advantageous when the relationship between y and x is non-linear (Sukhatme & Sukhatme, 1984).

DOI: https://doi.org/10.57233/ijsgs.v10i2.651 ISSNp: 2488-9229; ISSNe: 3027-1118

Various modifications and extensions exist, each with specific advantages and drawbacks depending on the sampling design and population characteristics (Khare & Srivastava, 2012). Choosing between separate and combined estimators involves understanding their trade-offs. While combined estimators often show higher efficiency in specific scenarios, they can be more complex to implement and require stronger assumptions about the relationship between variables. Additionally, their performance depends on the quality and relevance of the auxiliary information.

The field of survey sampling continues to evolve, with ongoing research exploring new combined estimators, incorporating complex sample designs, and handling missing data. Machine learning techniques are also being investigated for their potential to enhance estimation accuracy. Separate and combined ratio and regression estimators offer valuable tools for improving survey estimates. Understanding their strengths, weaknesses, and comparisons in the literature is crucial for informed selection and application in diverse sampling contexts. In the realm of survey research, a critical consideration is the presence of information pertaining to every unit within the population under investigation. Occasionally, this information may not pertain directly to the variable of interest but serves a crucial role in refining the sampling strategy or augmenting the precision of estimates for the variable of interest. Such auxiliary information is referred to as an auxiliary variable, providing supplementary insights into the characteristics of the variable under scrutiny (Cochran, 1977). Importantly, this auxiliary variable is typically known before data collection commences and is directly linked to the variable of interest (Särndal et al., 1992b).

In this research, the aim is to evaluate the efficiency of the ratio and regression estimators to determine the optimal approach. The objectives are to estimate separate and combined ratio estimators, estimate separate and combined regression estimator and compare separate and combined estimator under the two estimators (ratio and regression). This is a rare study with the target population in this study.

2.0 MATERIALS AND METHODS

2.1 Data Description

The data used for this study is secondary data collected from the Department of Statistics at Ladoke Akintola University of Technology on the scores of students in Probability Distribution III (STA302) and Statistical Inference II (STA411) along with their scores in Probability Distribution II (STA208) and Statistical Inference (STA204) which are their pre-requisite courses. The data was collected for 2021/2022 session. Microsoft Excel and RStudio were used for the statistical analysis.

2.2 Determination of Sample Size

Several studies have used sample size to select the number of samples needed (Althubaiti, 2023; Aslam et al., 2023; Constantin et al., 2023; de Bekker-Grob et al., 2015; Ko & Lim, 2021; Rajput et al., 2023; Salari et al., 2023; Sim et al., 2022). The Cochran correction method is used to obtain the sample size from the population if the population in finite.

$n = \frac{Z^2pq}{e^2}$ (Bartlett et al., 2001; Cochran, 1977) (1)

where; e is the desired level of precision (i.e. the margin of error = 5%).

p is the (estimated) proportion of the population which has the attribute in question (compuyed from the data).

q is $1-p$.

The Z value is found in z-statistical table $(\alpha = 0.01)$.

2.3 Ratio Estimation Method

Let a simple random sample of *n* farming households be selected without replacement from *N* such households in the community. Using the sample, the estimates $\tilde{y}$ of the mean yam sales and x of the mean cassava sales are obtained. These estimates are then used to calculate the required ratio using the formula.

$$\hat{R} = \bar{y}/\bar{x} = \sum_{i=1}^{n} y_i \,/\, \sum_{i=1}^{n} x_i \quad (2)$$

The ratio estimator $\hat{R}$ unlike its components $\bar{y}$ and $\bar{x}$ is biased. To obtain the bias of $\hat{R}$,

$$\text{Let } \bar{x} = \bar{X} + \bar{x} - \bar{X} = \bar{X}(1 + \overline{\alpha x}); \ \overline{\alpha x} = (\bar{x} - \bar{X})/\bar{X},$$

$$B(\hat{R}) = E(\hat{R}) - R = E(\hat{R} - R) = E\left(\frac{\bar{y} - R\bar{x}}{\bar{x}}\right)$$

$$E(\hat{R} - R) = \frac{E(\bar{y} - R\bar{x}(1 + \overline{\alpha x}^{-1}}{\bar{X}}$$

$$\hat{R}\text{-}R = \frac{(\bar{y} - R\bar{x})}{\bar{X}}(1 - \overline{\alpha x} + \delta^2\ \bar{x} - \delta^1\ \bar{x} + \ldots)$$

$$\hat{R}\text{-}R = \frac{(\bar{y} - R\bar{x})}{\bar{X}}(1 - \overline{\alpha x}) = \frac{\bar{y} - R\bar{x}}{\bar{X}} - \frac{(\bar{y} - R\bar{x})\overline{\alpha x}}{\bar{X}}$$

$$*B(\hat{R}) = \frac{E\,(\bar{y} - R\,\bar{x}\,)\,\overline{\alpha x}}{\bar{X}} = \frac{E\,(\bar{y} - R\,\bar{x}\,(\bar{x} + \bar{X}\,)}{\bar{X}^2}$$

$$B(\hat{R}) = -\frac{1}{\bar{X}^2}\,[E\,\{\bar{y}\,(\,\bar{x} - \bar{X}\}] - E\,\{R\bar{x}\,(\,\bar{x} - \bar{X})\}$$

$$= -\frac{1-\int}{n\bar{X}^2}\,[S_{xy} - RS_x^2 - S_{xy}\,] \qquad (3)$$

where $S_{xy} = \sum_{|n|}^{N}(X_1 - \bar{X})\,(Y_1 - \bar{Y})\,/(N-1)$ is the covariance between y. S_x^2 is the population variance of the variable x.

2.3 Regression Estimation of Population Mean

Let $y_h x_i (I = 1, 2, ..., n)$ be the values of the main character y and auxiliary character *x* respectively obtained on a simple random sample of size *n* selected without replacement from a population of size *N*. as estimator of the population mean of *y* is:

$$\bar{y}_d = \bar{y} - \text{B}(\bar{x} - \bar{X}) \qquad (4)$$

B is a suitably chosen constant.

The estimator $\bar{y}_d$ is called difference estimator, which is unbiased for $\bar{Y}$.

The variance is :

$$V(\bar{y}_d) = V\,(\bar{y}\,) + \text{B}^2 V(\bar{x}) - 2\text{BCov}(\bar{x}, \bar{y})$$

$$= \frac{1-\int}{n}\,[S_y^2 + B^2 2BS_{xy}\,] \qquad (5)$$

3.0 Results and Discussion

In this research, the authors evaluated the use of This study investigated the use of separate and combined stratified random sampling to estimate population mean scores for two important courses in Statistics using their Pre-requisites as auxiliary information (Probability Distribution and Statistical Inference) for undergraduate students. Ratio and regression estimators were used to determine the optimal method. The data from the Department of Statistics at Ladoke Akintola University of Technology for the 2021/2022 academic session was used. These techniques stipulate a significant tool to appraise the consequences and expectations between the pre-requisite courses and the other courses, using the pre-requisite courses as auxiliary information. Various calculations were performed, including the sample size determination. Table 1 shows the summary statistics for Probability Distribution III (STA302) and Statistical Inference II (STA411) along with their scores in Probability Distribution II (STA208) and Statistical Inference (STA204) which are their pre-requisite courses. Overall, STA 204 has the highest mean of 49.14 when compared to other courses, and STA 302 has the lowest mean of 29.69 when compared to other courses. STA 411 has the highest standard deviation of 16.3806 when compared to other courses, this indicates that STA411 course has scores with higher variability.

Table 1: Descriptive Statistics

	STA302	STA208	STA411	STA204
Mean	29.69	48.45	33.95	49.14
Median	24.00	46.00	31.00	50.00
Maximum	85.00	88.00	75.00	81.00
Minimum	9.00	14.00	11.00	11.00
Observation	137	137	123	123
Standard Deviation	15.6114	16.3819	16.3806	15.0166

Table 2 presents the estimates of the sample mean, variance, covariance, and correlation coefficient of the scores of students in STA 411 (study variable) and STA 204 (pre-requisite course scores/auxiliary variable). The pre-requisite course scores/auxiliary variables STA 208 and STA 204 have a higher mean of 51.97 and 49.90 respectively, when compared to the mean of STA 302 and STA 411 respectively. This means that the students' performance was better in the pre-requisite courses. The P-value (0.2816) for the probability courses means that there is no evidence of a linear relationship between STA 302 and STA 208, because the p-value is greater than the level of significance $\alpha = 0.05$. The or the correlation coefficient (0.3500) means that the correlation between STA302 and STA411 is weak, though in a positive direction.

The P-value (0.0214) for the probability courses means that there is evidence of linear relationship between STA 411 and STA 204, because the p-value is less than the level of significance $\alpha = 0.05$.. The correlation coefficient (0.7191) means that the correlation between STA411 and STA204 is strong, and it is in a positive direction.

DOI: **https://doi.org/10.57233/ijsgs.v10i2.651** ISSNp: 2488-9229; ISSNe: 3027-1118

IJSGS FUGUSAU VOL10 (2) **WEBSITE: https://fugus-ijsgs.com.ng**

Table 2: Summary Statistics from the Sample Scores Selected Probability Distribution and Statistical Inference

	STA302 (y)	STA208 (x)	STA411 (y)	STA204 (x)
Mean	37.34	51.97	33.30	49.90
Variance	848.41	270.03	332.98	206.78
Co-variance		168.8		188.69
Correlation		0.3500		0.7191
P-value		0.2816		0.0214

Table 3: Mean scores and Variances using Ratio estimator for Statistical Inference and Probability Distribution

	STA411	STA302
Mean $\bar{y}_r$	32.79	34.81
Variance $\bar{y}_r$	0.9987	1.8792
Ratio estimation r	0.6673	0.7185

Table 3 shows the mean, variance and ratio estimation using the ratio method. Table 4 shows the slope, mean and variance using regression estimate. The regression method is a procedure of obtaining estimate of the population mean score by regressing the study variable on the auxiliary variable with the assumption that the study variable is significantly linearly related to the auxiliary variable.

Table 4: Mean scores and Variances using Regression estimators for Statistical Inference and Probability Distribution

	STA411	STA302
Slope β	0.9125	0.6251
Mean $\bar{y}_{reg}$	32.41	35.14
Variance $\overline{\text{var}}_{reg}$	166.55	765.39

The estimates of the mean score using the ratio and regression estimator have already been obtained for Statistical Inference (STA 411) and Probability Distribution (STA 302), the best estimates is then selected by comparing their variances, the best estimate is the one with the minimum variance. The smaller the variance of an estimate the better the estimate (De Menezes et al., 2021; Ibidoja et al., 2023a, 2023b; Jenkins & Quintana-Ascencio, 2020; Ravi et al., 2024).

The ratio estimation with a variance of 0.9987 for STA 411 and a variance of (1.8792) for STA 302 is the best estimator selected to obtain the mean score when compared to the regression estimation

Table 5: Mean and Variance for Separate ratio estimator

	STA411	STA302
$\bar{y}_{rs}$	34.4546	31.6061
Var(y_{rs})	113.5561	6612.1824

Table 5 shows the mean and variance using a separate ratio estimator. This employs first the stratified mean of the Y's and the X's and then defines the combined ratio estimator.

Table 6: Mean and Variance for combined ratio estimator

	y_{st}	x_{st}	r_0	y_{cr}	Var(y_{cr})
STA411	33.3	49.9	0.6673	33.9931	84.5372
STA302	32.6	51.95	0.6275	31.6589	6913.1749

DOI: https://doi.org/10.57233/ijsgs.v10i2.651 ISSNp: 2488-9229; ISSNe: 3027-1118
IJSGS FUGUSAU VOL10 (2) WEBSITE: https://fugus-ijsgs.com.ng

From Table 5 and Table 6 , the combined ratio estimate, the variance for the estimated variance score for STA 411 (84.5372) and for separate ratio estimate the variance is (113.5561). The better estimate for the mean score STA411 is the Combined ratio estimate because it has the smaller variance when compared to the Separate ratio estimate. While for STA 302, the combined ratio estimate is 31.6589 with variance (6913.1749) and the separate ratio estimate is (31.6061) with variance (6612.1824); hence the separate ratio estimate is the better estimate compared to the combined ratio estimate because it has the smaller variance.Thus, the combined ratio estimator is the best estimator for estimating the mean score in STA 411 while the Separate ratio estimator is the best for STA 302. From Table 7 and Table 8, the variance of the separate regression estimate is (165.0716) and combined regression estimate has variance of (17.5881), for STA411, following the condition of minimum variance, Combined regression estimate is the better estimate compared to the separate regression estimate. While for STA302, the variance of the combined regression estimate is (2220.2234) and separate regression estimate has a variance of (7.2189), the better estimate is the Separate regression estimate due to its smaller variance. Thus, separate regression estimator is the best for estimating the mean score in Probability Distribution (STA302), while Combined regression estimator is the best for estimating Statistical Inference (STA 411).

Table 7: Mean and variance for separate regression estimator

	y_{reg}	y_{sreg}	$Var(y_{reg})$
STA411	32.61	34.2433	165.0716
STA302	35.14	32.1843	7.2189

Table 8: Mean and Variance for Combined regression estimator

	y_{st}	x_{st}	β_0	y_{creg}	Var(y_{creg})
STA411	33.3	49.9		36.2368	17.5881
STA302	32.6	51.95	0.0313	32.992	2220.2234

4.0 CONCLUSION AND FUTURE WORK

In this research, the authors evaluated the efficiency of the ratio and regression estimators to determine the optimal approach. The objectives are to estimate separate and combined ratio estimators, estimate separate and combined regression estimators, and compare separate and combined estimators under the two estimators (ratio and regression). The limitation of the study is that it is difficult to find a recent study in the literature, where a similar method was used for the comparison. Another limitation is the accuracy of the data collected, this will affect the results. A separate regression estimator is best for estimating the Probability Distribution (STA302) while combined regression estimator is best for estimating Statistical Inference (STA 411).

For further studies, the effects of interaction and gender performance can be incorporated into the model, because it is crucial to understand the effect of male and male on the performance of the models. A hybrid model can also be developed for better estimates.

DECLARATION OF COMPETING INTEREST
There is no conflict of interest among the authors.

DATA AVAILABILITY
All data are available on request from the corresponding author.

REFERENCES

Ahmad, S., Ullah, K., Zahid, E., Shabbir, J., Aamir, M., Alshanbari, H. M., & El-Bagoury, A. A. A. H. (2023). A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-30150-9

Ahmad, S., Zahid, E., Shabbir, J., Aamir, M., Emam, W., Tashkandy, Y., & Ahmad, S. (2023). A simulation study: An enhanced generalized class of estimators for estimation of population proportion using twofold auxiliary attribute. Heliyon, 9(6). https://doi.org/10.1016/j.heliyon.2023.e17269

Ahmad, S., Zahid, E., Shabbir, J., Aamir, M., & Onyango, R. (2023). Enhanced Estimation of the Population Mean Using Two Auxiliary Variables under Probability Proportional to Size Sampling. Mathematical Problems in Engineering, 2023, 1–14. https://doi.org/10.1155/2023/5564360

Althubaiti, A. (2023). Sample size determination: A practical guide for health researchers. In Journal of General and Family Medicine (Vol. 24, Issue 2, pp. 72–78). John Wiley and Sons Inc. https://doi.org/10.1002/jgf2.600

Alyasin, A., Nasser, R., El Hajj, M., & Harb, H. (2023). Alyasin, A., Nasser, Assessing Learning Outcomes in Higher Education: From Practice to Systematization. TEM Journal, 12(3).

Aslam, I., Noor-ul-Amin, M., Hanif, M., & Sharma, P. (2023). Memory type ratio and product estimators under ranked-based sampling schemes. Communications in Statistics - Theory and Methods, 52(4), 1155–1177. https://doi.org/10.1080/03610926.2021.1924784

Balogun, O. S., Olaleye, S. A., & Ibidoja, O. J. (2020, November). University student's academic performance: an approach of Tau statistic. Proceedings of the 36th International Business Information Management Association (IBIMA).

Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational Research: Determining Organizational Research: Determining Appropriate Sample Size in Survey Research Appropriate Sample Size in Survey Research. Information Technology, Learning, and Performance Journal, 19(1).

Bee, M., Benedetti, R., Espa, G., & Piersimoni, F. (2010). On the use of auxiliary variables in agricultural survey design. Agricultural Survey Methods, 107–132.

Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.

Constantin, M. A., Schuurman, N. K., & Vermunt, J. K. (2023). A General Monte Carlo Method for Sample Size Analysis in the Context of Network Models. Psychological Methods. https://doi.org/10.1037/met0000555

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. Patient, 8(5), 373–384. https://doi.org/10.1007/s40271-015-0118-z

De Menezes, D. Q. F., Prata, D. M., Secchi, A. R., & Pinto, J. C. (2021). A review on robust M-estimators for regression analysis. In Computers and Chemical Engineering (Vol. 147). Elsevier Ltd. https://doi.org/10.1016/j.compchemeng.2021.107254

Hands, C., & Limniou, M. (2023). Why Science Qualifications Should be a Pre-requisite for Psychology Degree Programs–A Case Study Analysis From a UK University. In High Impact Practices in Higher Education: International Perspectives . Emerald Publishing Limited., 51, 127–147.

Ibidoja, O. J., Ajare, E. O., & Jolayemi, E. T. (2016). Reliability Measures of Academic Performance. IJSGS, 2(4), 59–64. www.naij.com/649525-

Ibidoja, O. J., Shan, F. P., Sulaiman, J., & Ali, M. K. M. (2023a). Robust M-Estimators and Machine Learning Algorithms for Improving the Predictive Accuracy of Seaweed Contaminated Big Data. J. Nig. Soc. Phys. Sci, 5, 1137. https://doi.org/10.46481/jnsps.2022.1137

Ibidoja, O. J., Shan, F. P., Sulaiman, J., & Ali, M. K. M. (2023b). Robust M-Estimators and Machine Learning Algorithms for Improving the Predictive Accuracy of Seaweed Contaminated Big Data. J. Nig. Soc. Phys. Sci, 5, 1137. https://doi.org/10.46481/jnsps.2022.1137

Islam, F., Khan, S., Wilson, I., & Gooch, R. (2008). The Value of Pre-requisite Courses for Statistics. In Journal of Business Inquiry.

Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. PLoS ONE, 15(2). https://doi.org/10.1371/journal.pone.0229345

Khare, B. B., & Srivastava, S. K. (2012). Modified ratio-regression estimator under double sampling. Journal of the Indian Society of Agricultural Statistics, 66(3), 399–404.

Ko, M. J., & Lim, C. Y. (2021). General considerations for sample size estimation in animal study. Korean Journal of Anesthesiology, 74(1), 23–29. https://doi.org/10.4097/kja.20662

Kumar, S., & Choudhary, M. (2023). An Improved Estimator of Finite Population Variance Using two Auxiliary Variable SRS. Revista Colombiana de Estadistica, 46(1), 81–91. https://doi.org/10.15446/rce.v46n1.104427

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

López, M. (2023). The effect of sampling mode on response rate and bias in elite surveys. Quality and Quantity, 57(2), 1303–1319. https://doi.org/10.1007/s11135-022-01406-9

Nair, K. R. (1988). On ratio-cum-regression estimation in finite population sampling. Journal of Official Statistics, 4(4), 305–315.

Pal, S. K., & Singh, H. P. (2022). An efficient new approach for estimating the general parameter using auxiliary variable in sample surveys. Afrika Matematika, 33(3).

Rach, S., & Ufer, S. (2020). Which Prior Mathematical Knowledge Is Necessary for Study Success in the University Study Entrance Phase? Results on a New Model of Knowledge Levels Based on a Reanalysis of Data from Existing Studies. International Journal of Research in

Undergraduate Mathematics Education, 6(3), 375–403. https://doi.org/10.1007/s40753-020-00112-x

Rajput, D., Wang, W. J., & Chen, C. C. (2023). Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics, 24(1). https://doi.org/10.1186/s12859-023-05156-9

Ravi, K. P., Ali, M. K. M., & Ibidoja, O. J. (2024). Identifying heterogeneity for increasing the prediction accuracy of machine learning models. Journal of the Nigerian Society of Physical Sciences, 2058. https://doi.org/10.46481/jnsps.2024.2058

Salari, N., Morddarvanjoghi, F., Abdolmaleki, A., Rasoulpoor, S., Khaleghi, A. A., Hezarkhani, L. A., Shohaimi, S., & Mohammadi, M. (2023). The global prevalence of myocardial infarction: a systematic review and meta-analysis. BMC Cardiovascular Disorders, 23(1). https://doi.org/10.1186/s12872-023-03231-w

Särndal, C.-E., Swensson, B., & Wretman, J. (1992a). Model-assisted survey sampling. Springer- Verlag.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992b). Model-assisted survey sampling. Springer- Verlag.

Shaffer, J., Schriner, S., Loudon, C., Dacanay, S., Alam, U., Dang, J., Aguilar-Roca, N., Kadandale, P., & Sato, B. (2018). The Impact of Physiology Pre-requisites on Future Anatomy and Physiology Courses. HAPS Educator, 22(3), 199–207. https://doi.org/10.21692/haps.2018.025

Sim, M., Kim, S. Y., & Suh, Y. (2022). Sample Size Requirements for Simple and Complex Mediation Models. Educational and Psychological Measurement, 82(1), 76–106. https://doi.org/10.1177/00131644211003261

Srivastava, S. K. (1967). AN ESTIMATOR USING AUXILIARY INFORMATION IN SAMPLE SERVEYS. Calcutta Statistical Association Bulletin, 16(2), 121–132.

Subzar, M., Ahmad Lone, S., Aslam, M., Hussein AL-Marshadi, A., & Maqbool, S. (2023). Exponential ratio estimator of the median: An alternative to the regression estimator of the median under stratified sampling. Journal of King Saud University - Science, 35(3). https://doi.org/10.1016/j.jksus.2022.102536

Sukhatme, P. V. , & Sukhatme, S. (1984). Sampling theory of surveys with applications. Iowa State University Press.

Theobald, C. M. (2014). A primer on regression for survey data. SAGE Publications.