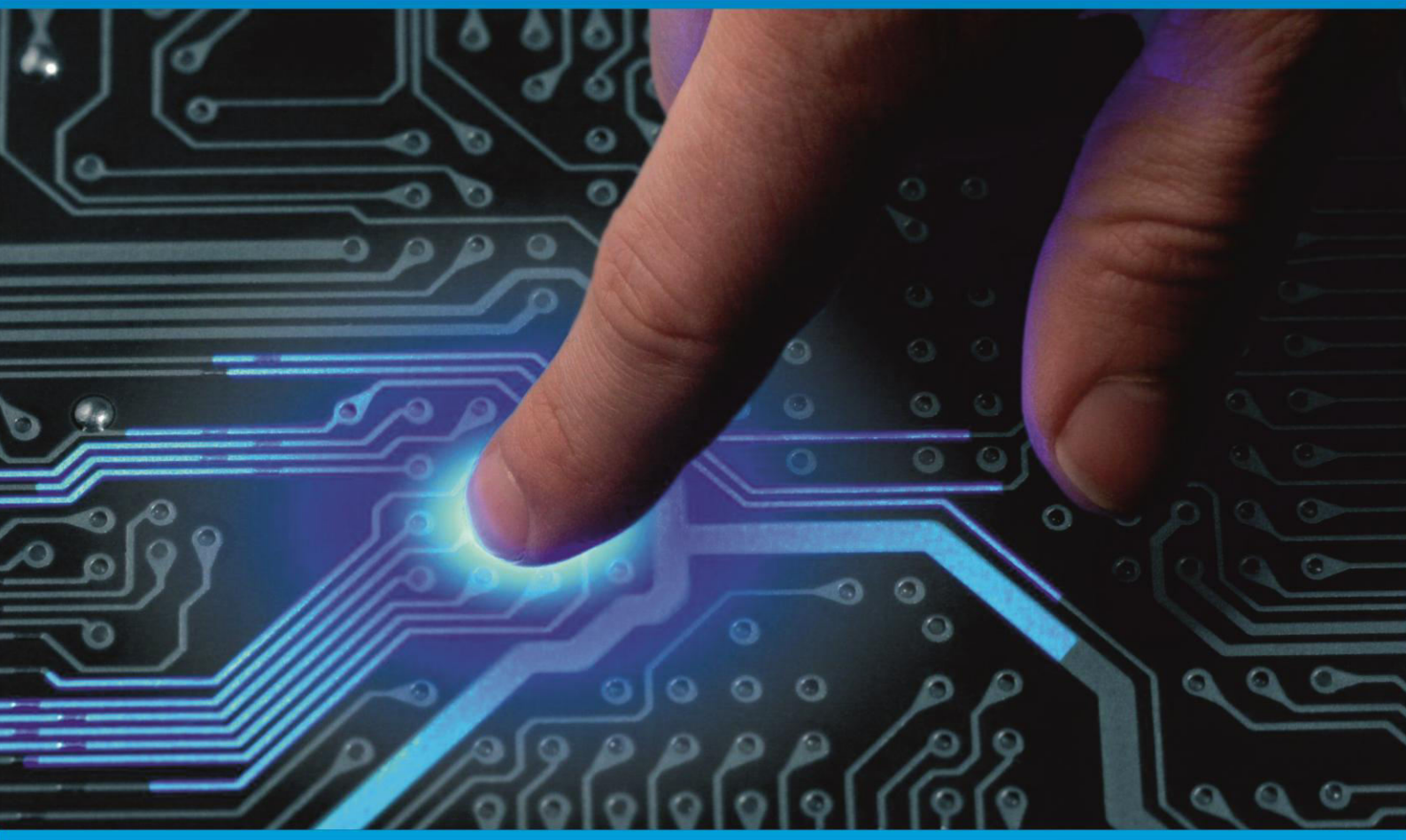
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

An Analysis on Different Levels of Education

Arushi Golia[1], Siddharth Nanda[2]

U.G Student, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [1]

Faculty, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [2]

ABSTRACT: This research paper in on Stratified Random Sampling on the dataset selected. The dataset had two variations, female and male and according to that, how well thay are performing in different areas like college, school, bachelor's degree or master's degree using marks in different fields like maths, reading and writing. Using this data I will predict the performance of male and female students based on thier level of education.

KEYWORDS: performance, stratified, sampling, student**,** level

I. INTRODUCTION

In this paper, we have using some concepts like Simple Random Sampling and Stratified Random Sampling. In Simple Random Sampling each sample has an equal probability to be picked which represents that it is unbiased in terms of population sample. It is used to represent the entire population data. It is the easiest method to extract a sample data from population data. In Stratified Random Sampling, we divide the data into homogeneous sub-groups known as strata. Random samples are then selected from each of these stratums. Stratified sampling provides better inclusion of the population. Using these methods we will find out the performance of different students categorized by gender and degree.As my dataset can be divided into groups, I used stratified sampling so that every group can participate equally into the result.
I have selected a dataset of 1000 students in different education levels like school, high school, bachelor's degree etc. the data is then categorized into math scores by every student which tells the average score one should have for a specific level of education.

II. LITERATURE SURVEY

[1] Academic performance and behavioral patterns.
ValentinKassarnig, EnysMones, Andreas Bjerre-Nielsen, PiotrSapiezynski, David Dreyer Lassen &Sune Lehmann tells us the impact of performance on individual and group of students and identifies the ways high performers are different from low performers and vice versa within a dataset of 538 entries.

[2] Student's Academic Performance: The Role of Motivation, Strategies, and Perceived Factors Hindering Liberian Junior and Senior High School Students Learning.
Charles Gbollie and Harriett Pearl Keamutell about the life of students and hindrances faced by them during schooling. How academics plays an important role in schooling and how marks varyas we go from lower level of education to higher level of education.

[3] Factors Affecting Students' Academic Performance.
IrfanMushtaq&Shabana Nawaz Khan tells us the important factors affecting the students' academic performance. There are various factors that affect a student's performance like family stress, poverty, proper guidance and many more.

[4] Antecedent Factors Affecting the Academic Performance of Graduate Students at the Nairobi Evangelical Graduate School of Theology.
Rosemary WahuMbogo tells us the positive and the negative impacts on a student in Africa affecting their academic performance and how to convert those negative impressions into positive ones.

[5] Factors Influencing Postgraduate Students' Performance: A high order top down structural equation modelling approach.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

Noor UlHadi, Barudin Muhammadtalks about the post-graduation of a student and how he/she develops a curiosity to learn more and research. Their main priority is to study. Also, these group of students contribute most into the sectors.

[6] A STUDY ON ACADEMIC PERFORMANCE OF UNIVERSITY STUDENTS.
ErumShahzadi, Z. Ahmad emphasizes the importance of faculty-student relation outside the classroom and clearing more doubts and asking for more study material helps student gain more knowledge than a normal student who just sits in the class.

[7] Stratified Sampling.
Van L. Parsons tells us about stratified sampling, its uses, its advantages, disadvantages and how it is better than simple and cluster sampling. Also this technique is used to increase efficiency of a sample as well as it is used in the modern technology a lot.

[8] Study on a Stratified Sampling Investigation Method for Resident Travel and the Sampling Rate.
YongjunShenincludes some surveys done in two cities using stratified random sampling and it tells the characteristics of travelling in both the cities.

[9] Stratified-sampling over social networks using MapReduce.
Roy Levin, YaronKanzatalks about the advantages on Stratified Sampling and how to implement it using MapReduce. It also talks about the problem of stratified sampling for selecting sample population on large scale.

[10] Stratified Sampling of Neighborhood Sections for Population Estimation.
Roger Hillson, Joel D. Alejandre, Kathryn H. Jacobsen, Rashid Ansumana, Alfred S. Bockarie, UmaruBangura, Joseph M. Lamin, and David A. Stengercreated two models in this paper and stratified sampling was sued in both the cases. Cluster sampling is also explained and compared with stratified sampling.

III. PROPOSED ANALYSIS APPROACH

Asurvey was conducted to see the number of students opting different levels of education using their math score. A sample size of 278 students was taken out of 1000 students for 6 different levels of education as shown below:

Levels of Education	Sample Size	Math Score (Sum)
Associate	47	3023
Bachelor's	46	3239
Master's	46	3229
High School	46	2722
Some College	47	3101
Some High School	46	2804

Using sample data, we need to estimate the mean math score in the population. Also need to find the margin of error and the confidence interval at 95% confidence interval.

1. Population Parameter:

To calculate the overall sample mean, we first need to compute the sample mean of each stratum.

$\bar{x} = \sum(x_i)/n$

Where, n = sample size,
x_i = sample math sum

Mean (Associate) = $\bar{x}_{associate} = \frac{3023}{47} = 64.32$

Mean(Bachelor's) = $\bar{x}_{bach} = \frac{3239}{46} = 70.41$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

Mean (Master's) = $\bar{x}_{mast} = \frac{3229}{46} = 70.2$

Mean (High School) = $\bar{x}_{HS} = \frac{2722}{46} = 59.17$

Mean (Some College) = $\bar{x}_{SC} = \frac{3101}{47} = 65.97$

Mean(Some High School) = $\bar{x}_{SHS} = \frac{2804}{46} = 60.95$

After calculating mean of every stratum, the overall sample mean is:

Sample Mean = (64.32+70.41+70.2+59.17+65.97+60.95) / 6

$= \frac{391.04}{6} = 65.17$

Thus, based on the data from the sample strata, we estimate that the mean math score in the sample is equal to 65.17.

Population Mean = $(N_h/N)*x_h$

Where N_h = Actual Sample Size
N = Population Size
x_h = sample mean

$= \left(\frac{222}{1000} * 64.32\right) + \left(\frac{118}{1000} * 70.41\right) + \left(\frac{59}{1000} * 70.2\right) + \left(\frac{196}{1000} * 59.17\right) + \left(\frac{226}{1000} * 65.97 + (1761000 * 60.95)\right)$

$= 63.9$

Based on the data, we can estimate the mean math score for the population is equal to 63.9.

2. Population Variance

We need to compute sample variance within each stratum, to compute standard error for every stratum.

$s^2_{associate} = 231$

$s^2_{bach} = 215.89$

$s^2_{mast} = 205.98$

$s^2_{HS} = 334.65$

$s^2_{SC} = 242.71$

$s^2_{SHS} = 342.97$

3. Standard Error

The standard error estimate the variability of our sample estimate of the population mean. We will make use of standard error to determine the margin of error and to define a confidence level.

SE = (1 / N) * sqrt {Σ [N^2_h * (1 - n_h/N_h) * s^2_h / n_h] }

$SE_{associate} = 4.55$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

SE_{bach} = 4.46
SE_{mast} = 4.26
SE_{HS} = 5.47
SE_{SC} = 4.67
SE_{SHS} = 5.53

Thus, the standard error of the sampling distribution of the mean is 4.11.

We are working on 95% confidence level.

4. Confidence Level

There are two tests, t-test and z-test. T-test is used when we have a small sample size (<50) and when population variance is unknown whereas z-test is used when we have large sample size (>50) and population variance is known. As the sample size is large, thus we will use z-test. Standard Normal Distribution Table is used to find critical z-score. In this part of the analysis, usually researchers choose a confidence level and the most frequently chosen confidence level is 95%. Thus, we will use that only.

5. Critical Value

The critical value is used to figure out the margin of error.To calculate the critical value, first we need to find out:

- Alpha(α):
  α = 1-(confidence level/100)
  α = 1-(95/100)
  = 0.05
- Critical Probability (p^*) :
  p^*= 1- (α/2)
  = 1-(0.05/2)
  = 0.975

Using Standard Normal Distribution Table, we can see that the critical z-score value is 1.96.

6. Margin of Error

ME = (Critical Value * Standard Error)
= 1.96 * 4.11
= 8.05

7. Confidence Interval

The minimum and the maximum values of the confidence interval are:

CI_{min} = x – Standard Error * Critical Value
CI_{max} = x + Standard Error * Critical Value

CI_{min} = 65.17 – (4.11 * 1.96)
= 57.11
CI_{max} = 65.17 + (4.11 * 1.96)
= 73.22

Summary

Based on our sample data, we estimate that our population mean is 63.9. given a 95% confidence level, the margin of error around that estimate is 4; and the 95% confidence interval is from 57.11 to 73.22. Thus, we can see that the population mean is between the confidence interval which means an average of 63.9 math score is required for different levels of education.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

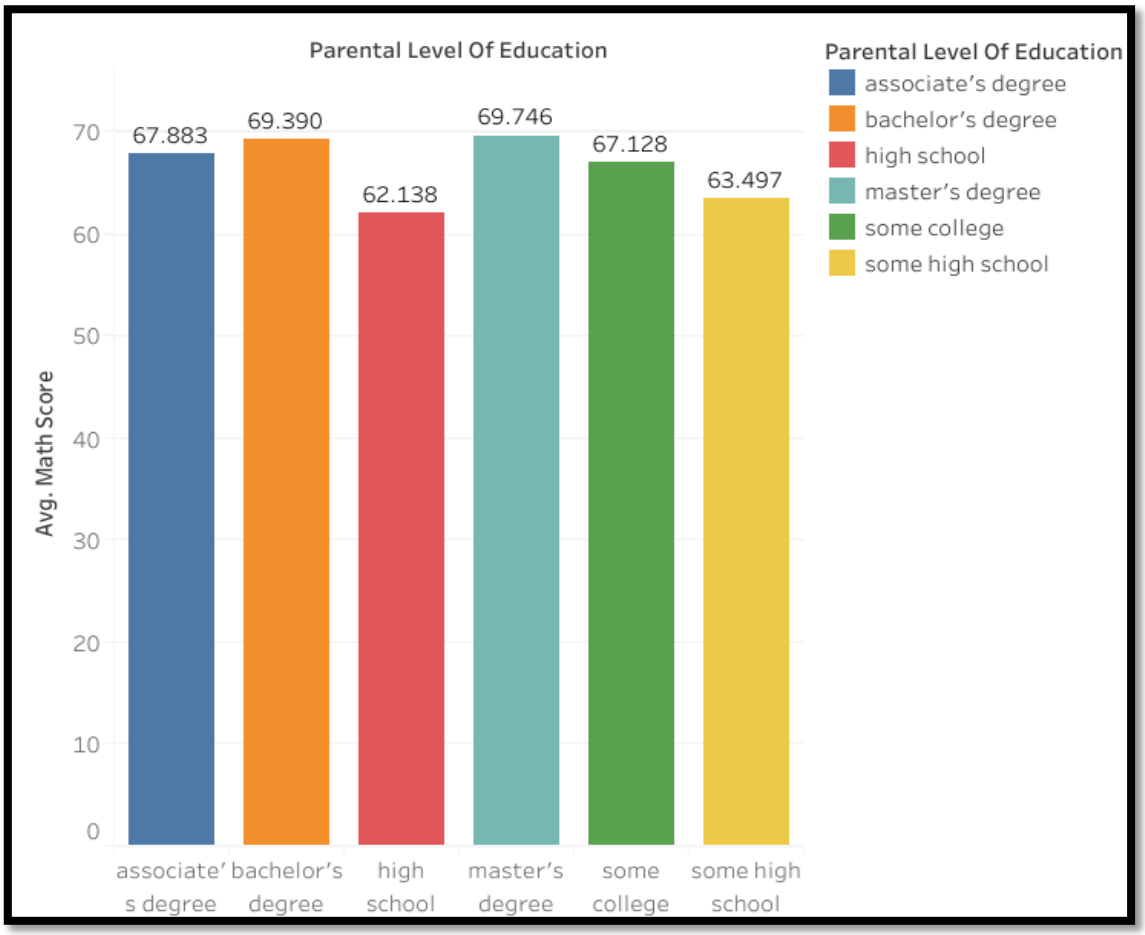


Fig 1: Level of Education V/S Avg. Math Score

IV. FUTURE SCOPE & DISCUSSION

I have used stratified sampling to categorize data so that all groups receive proper representation within the sample. Thus, it provides better provision of the population since it includes all the sub groups to see whether all are participating in the sampling.

This study focussed on math marks as the main parameter, we can increase this range to more areas like science or literature which are some common subjects. So if some students are not good in calculation but stand out in reading or writing skills, they can increase their mean through those subjects. We can also categorize this data into different streams/subjects opted by students in every level of education to have more specific idea in every field. We can also promote students to take more part in post-graduation studies as from the sample we can see that sample size of master's are much less than sample size of bachelor's.

V. CONCLUSION

In this research paper, we have taken a dataset consisting of different level of education and different scores of students categorized by levels of education. We have used Stratified Random Sampling to create some stratas within the dataset to have a uniform data. After creating those strata, we applied some statistics to it, to find the average value of math score needed for different levels of education as well as vaiance of every strata. The majority of students have an average above 60 when it comes to math scores. When calculated everything, the population mean comes between the confidence interval, which means the mean score which we calculated for our sample data is also applicable to population data. We can see from the graphs that the math score average is heighest in master's degree and lowest in highschool.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905042 |

REFERENCES

[1] Kassarnig, V., Mones, E., Bjerre-Nielsen, A., Sapiezynski, P., Lassen, D.D. and Lehmann, S., 2018. Academic performance and behavioral patterns. *EPJ Data Science*, *7*(1), p.10.
[2] Gbollie, C. and Keamu, H.P., 2017. Student academic performance: The role of motivation, strategies, and perceived factors hindering Liberian junior and senior high school students learning. *Education Research International*, *2017*.
[3] Mushtaq, I. and Khan, S.N., 2012. Factors affecting Studentsâ€™ academic performance.Global journal of management and business research, 12(9).
[4] Mbogo, R.W., 2016. Antecedent Factors Affecting Academic Performance of Graduate Students at the Nairobi Evangelical Graduate School of Theology.Journal of Education and Practice, 7(18), pp.128-141.
[5] Hadi, N.U. and Muhammad, B., 2019. Factors Influencing Postgraduate Students' Performance: A high order top down structural equation modelling approach. Educational Sciences: Theory & Practice, 19(2).
[6] Shahzadi, E. and Ahmad, Z., 2011. A study on academic performance of university students.Recent Advances in Statistics, 255.
[7] Parsons, V.L., 2014. Stratified sampling. Wiley StatsRef: Statistics Reference Online, pp.1-11.
[8] Shi, F., 2015. Study on a stratified sampling investigation method for resident travel and the sampling rate. Discrete Dynamics in Nature and Society, 2015.
[9] Levin, R. and Kanza, Y., 2014, June. Stratified-sampling over social networks using mapreduce.In Proceedings of the 2014 ACM SIGMOD international conference on Management of data (pp. 863-874).
[10] Hillson, R., Alejandre, J.D., Jacobsen, K.H., Ansumana, R., Bockarie, A.S., Bangura, U., Lamin, J.M. and Stenger, D.A., 2015. Stratified sampling of neighborhood sections for population estimation: A case study of Bo City, Sierra Leone.PloS one, 10(7), p.e0132850.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details