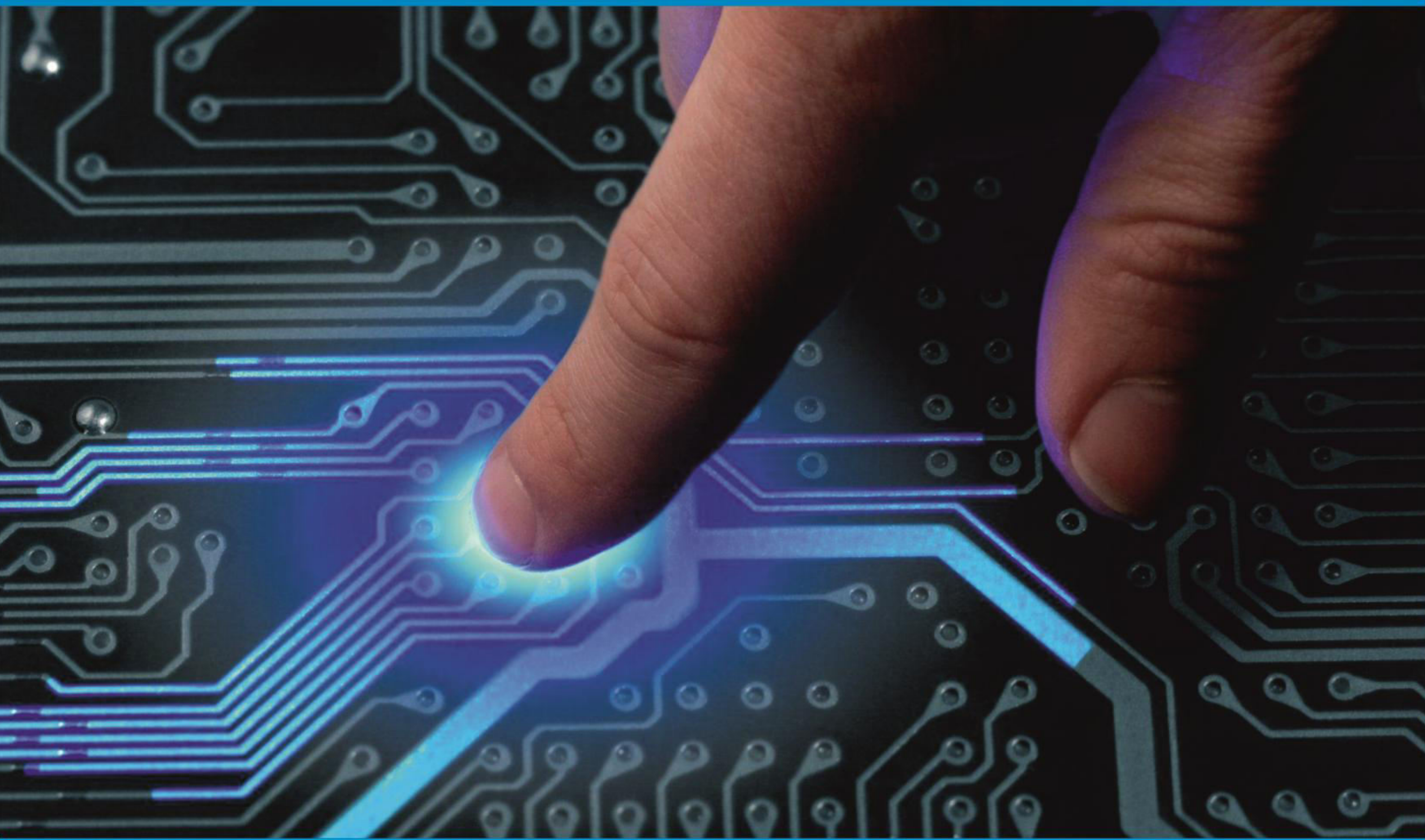
IJIRCCE

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

| 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.2021.0905043 |

Water Discharge Quality Prediction Using Stratified Sampling Method

Ataish Nehra[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: The observing of wastewater quality is fundamentally significant for the dependability of a biological system. In this study we use sampling method to predict on how one can identify the hazardous wastewater on the basis of the fecal coliform level and group them further into categories. This paper is a brief on how to use in particular stratified sampling method for finding the mean, confidence interval, margin of error, variance and standard error, while studying a large dataset for the wastewater quality prediction method. It also states some literature survey of a broader view of stratified sampling method as well as wastewater quality related field. It explains in detail all the factors and the mixture of technology with the aim of living a less hazardous environment by treating the sewer system.

KEYWORDS: Wastewater quality prediction, Machine learning, Stratified sampling, Artificial intelligence, Statistical analysis.

I. INTRODUCTION

To secure the climate and human wellbeing, treated wastewater release should be tested and checked in most created nations to guarantee release licenses are met. The assortment and investigation of treated wastewater effluents is tedious and expensive. Albeit the diurnal variety of flowrates and wastewater quality is a verifiable truth in wastewater designing, wastewater engineer's configuration plants that will create treated release that meets the grant cutoff points of water quality. One can conjecture that even in streams and lakes with different dischargers, because of the designing nature of the plants that produce the gushing, apparatuses could be utilized to estimate possible emanating quality. The effect of such devices would bring about the decrease in inspecting recurrence and minimization of expenses related with examination. Previously, researchers needed to gather and investigate countless wastewater tests to see how their parts affected the climate. At that point deterministic models and multivariate direct relapse (MLR) examination were utilized to accelerate the interaction of assessing the nature of wastewater gushing releases. As a water quality dataset is a sort of time arrangement dataset, which is probably going to have a muddled nonlinear relationship, the presentation of deterministic and MLR models is bad. Instead, stratified sampling is used here in order to find the estimate. Stratified sampling is a likelihood examining technique that segments the whole dataset into various subgroups, or layers, in light of the estimation of every information point. At that point scientists relatively select information focuses from various layers for various purposes. This technique is broadly utilized in characterization issues to check proposed models, in which the preparing and testing datasets of each overlay contain generally a similar extent of each class name. Contrasted and the conventional examining strategy, which produces n irregular parts from the first dataset, defined testing ensures the preparing and testing dataset can equitably cover the entirety of the various classifications. A defined inspecting strategy works better in situations where the issue has information sparsity limitations. When testing from an enormous dataset, haphazardly testing information from the gathered dataset can cover all situations. Notwithstanding, when it comes to a medium, or even a little significant degree dataset, a few sorts of situations most likely drop out of the inspected dataset. Since the majority of the water quality observing stations have just been implicit the previous quite a few years, and an enormous extent of the water quality boundaries have just been recordable in the previous 20 years, the water quality dataset has a place with this case, which has predetermined number of information tests.

ABOUT THE DATASET:
The data contains water quality parameters of different rivers of India. There are 8 parameters and each parameter is the average values measured over a period of time. The data has been combined from the official website for data related to India. It has been collected from Kaggle website as it is an open-source dataset.

| 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.2021.0905043 |

Average value of pH	Average value of conductivity	Average value of biochemical oxygen demand	Average value of nitrate-n and nitrite-n	Average value of fecal coliform
	NA 6%	1 3%	0 24%	NA 15%
	290 1%	1.5 3%	0.3 7%	4 5%
6.3 14.7	Other (498) 93%	Other (498) 93%	Other (367) 69%	Other (424) 79%
7.8	371	5.6	3.55	90
7.9	294	3.2	2.69	45
7.5	513	12.6	2.3	131
7.7	475	10.3	1.9	122
7.7	385	3.8	1	110
7.8	410	5.2	1.5	77

Fig 1: Glimpse of dataset.

II. LITERATURE SURVEY

1. **Effluent quality prediction in papermaking wastewater treatment processes using dynamic Bayesian networks.**

Compelling internet displaying of papermaking wastewater treatment measures (WWTPs) is a significant way to guarantee wastewater reusing and innocuous release. A composite model coordinating variable significance in projection with dynamic Bayesian organizations (VIP-DBN) is proposed to improve the displaying capacity and unwavering quality in WWTPs. Initial, a variable determination technique is applied to improve on the organization structure and decrease demonstrating costs. At that point the enlarged grids method is implanted into the Bayesian organizations to adapt to the unique qualities, nonlinearity, and vulnerability all the while. The demonstrating execution of VIP-DBN is assessed through two contextual analyses, a recreated WWTP dependent on benchmark reproduction model no. 1 (BSM1) and a genuine papermaking WWTPs, in which the VIP-DBN model shows preferable demonstrating execution over its adversaries. In particular, contrasted and halfway least squares, counterfeit neural organizations, and Bayesian organizations, the assurance coefficient estimation of VIP-DBN is expanded by 34.36%, 20.55%, and 3.30%, separately, for the forecast of profluent nitrate in BSM1. The VIP-DBN can be utilized to manage complex WWTPs in mechanical applications, which gives a superior viable technique to delicate sensor demonstrating and an assurance for the compelling dynamic of wastewater treatment measures for papermaking endeavors.

2. **Predictive modelling the discharge of urban wastewater using artificial intelligent models (case study: Kerman city).**

Metropolitan wastewater release is perhaps the main parts for the turn of events and plan of water and wastewater treatment projects. In this examination, the day-by-day metropolitan wastewater release (UWD) was anticipated utilizing two fake insightful models including the complex perceptron neural organization (MLPNN) and hereditary programming (GP). For this reason, for existing's, (Kerman city situated in the south-east of Iran) sewage information, which have been recorded day by day over the most recent 5 years, were utilized. To plan the information example of the models, up to four-time defers units were thought of. The consequences of the two models proclaimed that two defer units are sufficient for demonstrating and expectation of the UWD. The created MLPNN model comprises of two secret layers. The first and second secret layers have seven and five neurons, individually. The digression sigmoid was

| 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.2021.0905043 |

considered as move work overseeing the condition on neurons. The blunder records of created MLPNN in testing stages are R2 = 0.77 and RMSE = 1589. The design of the numerical model created dependent on hereditary programming has three genomes whose construction comprises of two units of time delay. The exactness of GP displaying is adequate; in any case, its accuracy is somewhat not exactly the MLPNN, yet its outcomes are more pragmatic.

3. **Requirements for integrated wastewater models — driven by receiving water objectives.**

The plan of proficient specialized measures for the decrease of water contamination necessitates that wastewater release guidelines are driven by accepting water targets. Be that as it may, such incorporated water quality administration is just conceivable when the effect on the amphibian biological system can be anticipated quantitatively by methods for coordinated wastewater models. Regularly, a couple of kinds of wastewater release impacts are applicable for the condition of the accepting water and, thusly, the construction of the model can be kept moderately straightforward when zeroing in on one of these effects. The method of issue situated model choice is delineated for three normal instances of intense water contamination, that is harmfulness from un-ionized smelling salts, clean danger from pathogenic miniature creatures and oxygen exhaustion.

4. **Evaluating statistical model performance in water quality prediction.**

Openness to sullied water while swimming or drifting or partaking in other sporting exercises can cause gastrointestinal and respiratory illness. It isn't exceptional for water bodies to encounter fast vacillations in water quality, and it is accordingly crucial to have the option to anticipate them precisely and on schedule to limit populace's openness to pathogenic living beings. E. coli is generally utilized as a pointer to quantify water quality in freshwater, and higher tallies of E. coli are related with expanded danger to ailment. For this situation study, we analyse the presentation of a wide scope of factual models in forecast of water quality by means of E. coli levels for the week-by-week information gathered over the mid-year months from 2006 to 2014 at the sporting site on the Oreti waterway in Wallace town, New Zealand. The models incorporate credulous model, numerous straight relapses, dynamic relapse, relapse tree, Markov chain, order tree, irregular backwoods, multinomial calculated relapse, discriminant examination and Bayesian organization. The outcomes show that Bayesian organization was better than the wide range of various models. In general, it had a forget about one and k-overlay cross approval blunder pace of 21%, while foreseeing most of cases of E. coli levels delegated hazardous by the Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas 2003, New Zealand. Since Bayesian organizations are likewise adaptable in dealing with missing information and exceptions and consider consistent refreshing continuously, we have discovered them to be a promising device, and later on, plan to expand the examination past the current contextual analysis site.

5. **Efficient estimation and stratified sampling.**

In this paper they explored the assessment of a class of semi-parametric models. The piece of the model that isn't indicated is the negligible dispersion of the logical factors. The examining is delineated on the reliant factors, inferring that the informative factors are not, at this point exogenous or auxiliary. They build up another assessor for this assessment issue and show that it accomplishes the semi-parametric effectiveness headed for this case. Furthermore, it showed that the assessor applies to various testing plans that have recently been dealt with independently.

6. **Refinement strategies for stratified sampling methods.**

In numerous PC analyses the ampleness of an offered test to give satisfactory factual assessments can't be resolved deduced, and consequently the capacity to broaden or refine a trial configuration might be significant. This paper portrays refinement procedures for the class of delineated exploratory plans, for example, Latin hypercubes, symmetrical clusters, and factorial plans. A couple of uses are given to exhibit their convenience.

7. **Optimized stratified sampling for approximate query processing.**

The capacity to around answer conglomeration questions precisely and productively is of incredible advantage for choice help and information mining instruments. Rather than past inspecting based examinations, it treats the issue as an improvement issue where, given a responsibility of inquiries, it also selects a separated arbitrary example of the first information with the end goal that the blunder in noting the responsibility questions utilizing the example is limited. A critical curiosity of our methodology is that it can tailor the selection of tests to be vigorous, in any event, for responsibilities that are "comparable" however not really indistinguishable from the given responsibility. At last, the procedures perceive the significance of considering the change in the information circulation in a principled way. It shows how our answer can be carried out on a data set framework, and present aftereffects of broad investigations on Microsoft SQL Server that exhibit the better nature of our technique thought about than past work.

| 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.2021.0905043 |

8. **Hybrid decision tree-based machine learning models for short-term water quality prediction.**

Water assets are an incredible establishment and financial turn of events, and are firmly identified with wellbeing and the climate. Precise expectation of water quality is the way to improving water the executives and contamination control. In this paper, two novel crossover choice tree-based AI models are proposed to get more precise transient water quality expectation results. The fundamental models of the two half breed models are outrageous angle boosting (XGBoost) and arbitrary woodland (RF), which individually present a high-level information denoising method - complete troupe exact mode decay with versatile commotion (CEEMDAN). Taking the water assets of Gales Creek site in Tualatin River (quite possibly the most contaminated streams on the planet) Basin for instance, a sum of 1875 information (hourly information) from May 1, 2019 to July 20, 2019 are gathered. Two crossover models are utilized to anticipate six water quality pointers, including water temperature, broken up oxygen, pH esteem, explicit conductance, turbidity, and fluorescent disintegrated natural matter. Six blunder measurements are presented as the premise of execution assessment, and the consequences of the two models are contrasted and the other four regular models. The outcomes uncover that: (1) CEEMDAN-RF performs best in the forecast of temperature, broken down oxygen and explicit conductance, the mean total rate blunders (MAPEs) are 0.69%, 1.05%, and 0.90%, separately. CEEMDAN-XGBoost performs best in the forecast of pH worth, turbidity, and fluorescent disintegrated natural matter, the MAPEs are 0.27%, 14.94%, and 1.59%, individually. (2) The normal MAPEs of CEEMDAN-RF and CEEMMDAN-XGBoost models are the littlest, which are 3.90% and 3.71% individually, demonstrating that their general forecast execution is the awesome. What's more, the solidness of the expectation model is likewise talked about in this paper. The examination shows that the forecast solidness of CEEMDAN-RF and CEEMDAN-XGBoost is higher than other benchmark models.

9. **Wastewater discharge impact on drinking water sources along the Yangtze River (China).**

Impromptu aberrant (accepted) wastewater reuse happens when wastewater is released into surface waters upstream of consumable drinking water treatment plant admissions. This paper expects to foresee rates and patterns of true reuse all through the Yangtze River watershed to comprehend the general commitment of wastewater releases into the waterway and its feeders towards deflecting water shortage concerns. The Yangtze River is the third longest on the planet and supports mutiple/15 of the total populace, yet the significance of wastewater on the stream stays not well characterized. Metropolitan wastewater created in the Yangtze River Basin expanded by 41% somewhere in the range of 1998 and 2014, from 2580 m3/s to 3646 m3/s. Under low stream conditions in the Yangtze River close to Shanghai, offered wastewater commitments waterway streams expanded from 8% in 1998 to 14% in 2014. The most significant levels of true reuse showed up along a significant feeder (Han River) of the Yangtze River, where true reuse can surpass 20%. While this underlying investigation of accepted reuse utilized water supply and wastewater information from 110 urban areas in the bowl and 11 measuring stations with > 50 years of memorable streamflow information, the result was restricted by the absence of checking stations at more areas (i.e., information must be anticipated utilizing advanced rise planning) and absence of exact geospatial area of drinking water admissions or wastewater releases. This restricted the prescient ability of the model comparative with bigger datasets accessible in different nations (e.g., USA). This evaluation is the primary examination of true wastewater reuse in the Yangtze River Basin. It will help distinguish segments of the waterway at higher danger for wastewater-related poisons because of quality of—and dependence on—wastewater release that could be the focal point of field studies and model expectations of higher spatial and worldly goal.

10. **Prediction analysis of a wastewater treatment system using a Bayesian network.**

Wastewater treatment is a confounded powerful interaction, the viability of which is influenced by microbial, substance, and actual elements. As of now, anticipating the emanating nature of wastewater treatment frameworks is troublesome on account of complex organic response instruments that shift with both time and the actual ascribes of the framework. Bayesian organizations are valuable for tending to vulnerabilities in man-made consciousness applications. Their incredible inferential ability and advantageous choice help instruments give adaptability and appropriateness to depicting and investigating factors influencing wastewater treatment frameworks. In this investigation, a Bayesian organization-based methodology for demonstrating and anticipating a wastewater treatment framework dependent on Modified Sequencing Batch Reactor (MSBR) was proposed. Utilizing the introduced approach, a Bayesian organization model for MSBR can be developed utilizing experiential data and actual information identifying with influent burdens, working conditions, and profluent fixations. Furthermore, MSBR expectation investigation, wherein emanating focus can be anticipated from influent burdens and operational conditions, can be performed. This methodology can be applied, with insignificant alterations, to different sorts of wastewater treatment plants.

| 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.2021.0905043 |

III. PROPOSED ANALYSIS APPROACH

- Estimate a population parameter.
- Estimate population variance.
- Compute standard error.
- Specify a confidence level.
- Find the critical value (often a z-score or a t-score).
- Compute margin of error.
- Define confidence interval.

Calculation for stratified sampling:

We are administering in the level of fecal coliform in wastewater through sample of drainage. Now according to the data set there are 20,000 drainages, half of which are for households and the other half for commercial wastage. Using hope proportionate stratified sampling to select 40 drainages for testing. As the population is half for household and half for commercial wastage, one stratum consists of 20 households and 20 commercial drainages.

Fecal coliform level four each sample is shown here.

H	131	122	120	110	100	122	122	122	130	130	130	128	128	128	122	122	110	110	110	112
C	122	122	122	120	120	120	130	132	132	132	132	124	124	124	124	134	134	116	116	118

For this stratified sampling procedure, we assume 95% confidence level.

Using sample data estimate the main reading achievement level in the population. We will find the margin of error and the confidence interval. Assuming a 95% confidence level.

To begin with our calculations, we follow the seven-step process for stratified sampling which are explained below along with the numerical calculation.

STEP1:

Estimating the population mean. To compute the overall sample mean, we need to compute the sample means for each stratum. The stratum mean for household is equal to:

X(household) = (Sum of all the sample elements in household sample/Total number of elements)

= (131+122+120+--------+110)/20 = 2409/20 = 120.45

The stratum mean for commercial drainage is computed similarly. It is equal to 124.9.

Therefore, overall sample mean is:

$$\bar{x} = \sum (N_H - N) * x_H$$

= (10,000/20,000)120.45 + (10,000/20,000)124.9

= 122.705

Therefore, based on data from the sample strata, we estimate that the main reading achievement level in the population is equal to 122.705.

STEP 2:

Now we will compute sample variance within strata. We need to compute the sample variance within each stratum, so we can compute the standard error in the next step. For household drainage the within stratum sample variance is equal to

$$S_H^2 = \frac{\sum (x_i - x_H)^2}{\eta_H - 1}$$

$= [(131 - 120.45)^2 + (122 - 120.45)^2$+----+$(112 - 120.45)^2]/(20-1)$

= 160.0734

The within stratum sample variance for commercial drainage is computed similarly. It is equal to 32.66.

| 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.2021.0905043 |

STEP 3:
In this step we will compute standard error. the standard error measures the variability of our sample estimate of the population mean. we will use standard error to compute the margin of error and to define our confidence level.

$$SE = (1 - N) * sqrt\{\varepsilon[N_H^2 * (1 - n_H N_H) * S_H^2 / n_H]\}$$

SE= (1/20000)*sqrt{[100000000*(1-20/10000)*160.0734/20]+[100000000*(1-20/10000)*32.66/20]}
SE= 1.553251477 or 1.55(approx.)

Hence, the standard error of the sampling distribution of the mean is 1.55.
We are working with a 95% confidence level.

STEP 4:
Now we will move on to finding the critical value. the critical value is a factor used to compute the margin of error. to find the critical value we take the steps:
Compute Alpha(α):
α= 1-(Confidence level/100)
= 1-(95/100) = 0.05
STEP 5:
Finding the critical probability (p*):
p*= 1-(α/2) = 0.975
NOTE: We use the t score when sample sizes are small; z score when it is large (at least 30). you can use the standard normal distribution table find the critical z score and t distribution table to find the critical t statistic.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

Fig2: Negative z table.

| 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.2021.0905043 |

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Fig3: Positive z table.

STEP 6:
Now we will compute the margin of error:
ME = critical value * standard error
ME = 1.96 * 1.55 = 3.038
STEP 7:
Specify the confidence interval. the minimum and maximum values of confidence interval are:
CI(min)= population mean- SE * CV = 122.705 – 3.038 = 119.667
CI(max)= population mean- SE * CV = 125.743

| 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.2021.0905043 |

IV. FUTURE SCOPE & DISCUSSION

Stratified sampling can be used through various programming languages for example it can be used through R-programming and that can be connected to a database which would provide live stream of constantly updated data based on waste water levels in various parts of a city. It would help in separating drainage lines according to the fecal coliform level and dangers that it possesses. A smart city should have all the technologies above and below the ground as well and by that it means that the sewer system should also be monitored through technology which is possible through this sampling method for a safer lifestyle.

Conclusion

In summary, here are the results of analysis based on sample data, we estimate that the population mean is 122.705. Given a 95% confidence level, the margin of error around the estimate is 1.55; find the 95% confidence interval is from 119.667 to 125.743.

REFERENCES

1. Phuong Thi Xuan Nguyen, Hien Van Hoang, HuyenThiKhanhDinh, Pierre Dorny, Bertrand Losson, Dung Thi Bui, Laetitia Lempereur, Insights on foodborne zoonotic trematodes in freshwater snails in North and Central Vietnam, Parasitology Research, 10.1007/s00436-020-07027-1, **120**, 3, (949-962), (2021).
2. Stefano Calzavara, Mauro Conti, Riccardo Focardi, AlviseRabitti, Gabriele Tolomei, undefined, 2019 IEEE European Symposium on Security and Privacy (EuroS&P), 10.1109/EuroSP.2019.00045, (528-543), (2019).
3. Ünal Aksu, Dennis M. M. Schunselaar, Hajo A. Reijers, Automated Prediction of Relevant Key Performance Indicators for Organizations, Business Information Systems, 10.1007/978-3-030-20485-3_22, (283-299), (2019).
4.] Chan SN, Thoe W, Lee JHW. Real-time forecasting of Hong Kong beach water quality by 3D deterministic model. Water Res 2013;47(4):1631–47.
5. French M, Recknagel F (1994) Modeling algal blooms in freshwaters using artificial neural networks. In: Zanetti P (ed) Computer Techniques in Environmental Studies V, vol II, Environment Systems. Computational Mechanics Publications, Boston, pp 87–94
6. Hatzikos E, Anastasakis L, Bassiliades N, Vlahavas I, (2005) Simultaneous prediction of multiple chemical parameters of river water quality with tide. In: Proceedings of the Second International Scientific Conference on Computer Science, IEEE Computer Society, Bulgarian Section.
7. Hem JD, (1985) Study and interpretation of the chemical characteristics of natural water, 3rd edn. US GeolSurv Water Supply Pap 2254, 263 p
8. Jang JSR (1993) ANFIS: Adaptive-network-based fuzzy inference systems. IEEE Trans Syst Man Cyb 23(3):665–685
9. Jang JS, Sun CT, Mizutani E, (1997) Neuro-fuzzy and soft computing. Prentice-Hall, Englewood Cliffs. ISBN 0-13-2874679.
10. Kim B, Park JH, Kim BS (2002) Fuzzy logic model of Langmuir probe discharge data. Comput Chem 26(6):573–581

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