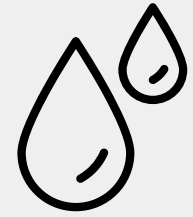# ANALYSIS OF THE RELATIONSHIP BETWEEN AMOUNT OF SILICON DIOXIDE AND PER CAPITA INCOME, ECONOMIC & POWER INEQUALITIES ACROSS DISTRICTS IN INDIA, 2018

HARSHIT CHOUDHARY (2021254)
KANISHK KUKREJA (2021393)
SHEFALI MITTAL (2019389)
YASHILA ARORA (2021436)

INTRODUCTION

Groundwater is an important natural resource. India being the largest user of groundwater in the world making it a vital resource and a vital aspect that should be investigated.

The concentration of SiO_2 , a naturally occurring substance present in groundwater in variable concentrations, can have a substantial effect on the social, economic, and environmental aspects of the communities that depend on this resource.
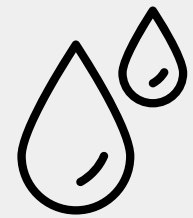
SiO_2 has many harmful impacts like Scaling in industrial process, corrosion of pipes and other infrastructure .

SiO_2 can build up in soil and streams, which can be harmful to aquatic life and ecosystems. Also Long-term exposure to silica dust may cause respiratory problems such as lung cancer, chronic obstructive pulmonary disease, or COPD, and silicosis.

VARIABLE DESCRIPTION

GINI
Measure of Distribution of income across the population of a particular district.

FACTORIES
No. of industries in a state.

SDP
Total worth of products and services generated inside the state borders during any given fiscal year.

POLITICAL PARTIES_ SEATS
Number of political parties present in a state per unit seat.

DEATH
No. of people who died in a particular state.

WIN MARGIN
Winning margin is a compressed variable which assigns weights to different margins and then takes an average. Therefore if on a scale of 1-100 the winning margin is 50, it implies that the average winning candidate won around 30-40% of the votes

GINI_STATE
Measure of Distribution of income across the population of a particular district.

Coefficients	Estiimate
Intercept(α_0)	1.016e+0
sdp	4.618e-05
sdp^2	-7.261e-11
sdp^3	2.449e-17
Gini	1.688e+00

Multiple R-squared	0.1066
p-value	< 2.2e-16

EQI = B0 + B1(SDP) + B2(SDP^2) + B3(SDP^3) + B4(GINI) + U

This model helps us understand how different factors relate to the amount of Silicon dioxide in a particular area. The coefficients show the relationship between each factor and the amount of Silicon dioxide. The intercept value of 10.16 indicates the predicted amount of Silicon dioxide when all factors are zero. A one-unit increase in SDP increases Silicon dioxide by 4.618e-05, but beyond a certain point, Silicon dioxide decreases as SDP increases as depicted by sdp^2 and sdp^3 variables. A one-unit increase in Gini increases Silicon dioxide by 1.688, but this relationship is not statistically significant because it has a p-value of

0.475 which is much greater than 0.05. The R-squared value of 0.1066 means that the model can explain 10.66% of the variation in Silicon dioxide, and p-value indicate that the model is statistically significant.

Further, Sum of Residuals came very close to zero. Thus, confirming that our model accounted for almost all the variation in the errors

Coefficients	Estimate
Intercept($\beta 0$)	3.800e+01
sdp	8.260e-06
sdp^2	-3.096e-11
sdp^3	1.253e-17
Gini	-6.655e+00
Politicalparties_seat	2.562e+00
Factories	-4.491e-04
Winning_Margin	-3.811e+00
Deaths	2.551e-05

Multiple R-Squared	0.7122
P-value	< 2.2e-16

EQI = B0 + B1(SDP) + B2(SDP^2) + B3(SDP^3) + B4(GINI) + B5(POLITICALPARTIES_SEAT) + B6(FACTORIES) + B7(WINNING_MARGIN) + B8(DEATHS) + U

This model shows how different factors relate to Silicon dioxide levels. The coefficients indicate the relationship between each regressor and Silicon dioxide present in groundwater. A one-unit increase in sdp increases Silicon dioxide, but beyond a point, Silicon dioxide decreases. The coefficient for Politicalparties_seat suggests that a one-unit increase is associated with a 2.562 increase in Silicon dioxide. On the other hand, a one-unit increase in Factories is linked to a decrease of 4.491e-04 in Silicon dioxide. Similarly, a one-unit increase in Winning_Margin is connected with a decrease of 3.811 in Silicon dioxide, while a one-unit increase in Deaths is associated with an increase of 2.551e-05 in Silicon dioxide. The model can explain 71.22% of the variation in Silicon dioxide, and the p-value show that the model is statistically significant.

MAXIMUM LIKELIHOOD ESTIMATION

Maximum Likelihood Estimation (MLE) results were found to be very similar to the regression model with only a very minute difference, which indicated that the regression model is a good fit for the data and the model parameters have been accurately estimated. Thus, it can be concluded that the regression model might be reliable and can be used for making predictions or inferences about the relationship between the independent and dependent variables.

```
Call:
lm(formula = (newdata_politicalparties$Amount.of.Silicon.dioxide) ~
    newdata_politicalparties$sdp + I(newdata_politicalparties$sdp^2) +
    I(newdata_politicalparties$sdp^3) + newdata_politicalparties$Gini +
    (newdata_politicalparties$Politicalparties_seat) + (newdata_politicalparties$Factories) +
    (newdata_politicalparties$Winning_Margin) + (newdata_politicalparties$Deaths),
    data = newdata_politicalparties)

Residuals:
    Min      1Q  Median      3Q     Max
-30.993  -1.830  -0.321   0.992 197.284

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                                  3.800e+01  1.437e+00  26.439  < 2e-16 ***
newdata_politicalparties$sdp                 8.260e-06  2.716e-06   3.041  0.00237 **
I(newdata_politicalparties$sdp^2)           -3.096e-11  3.143e-12  -9.850  < 2e-16 ***
I(newdata_politicalparties$sdp^3)            1.253e-17  1.165e-18  10.756  < 2e-16 ***
newdata_politicalparties$Gini               -6.655e+00  1.508e+00  -4.414 1.03e-05 ***
newdata_politicalparties$Politicalparties_seat 2.562e+00 3.285e-02  78.014  < 2e-16 ***
newdata_politicalparties$Factories          -4.491e-04  4.769e-05  -9.418  < 2e-16 ***
newdata_politicalparties$Winning_Margin     -3.811e+00  1.135e-01 -33.591  < 2e-16 ***
newdata_politicalparties$Deaths              2.551e-05  7.575e-07  33.680  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.877 on 5473 degrees of freedom
  (6187 observations deleted due to missingness)
Multiple R-squared:  0.7122,    Adjusted R-squared:  0.7118
F-statistic:  1693 on 8 and 5473 DF,  p-value: < 2.2e-16
```

```
> print(coef_lm)
               sdp               sdp_2               sdp_3            Gini Politicalparties_seat
      8.260241e-06       -3.096177e-11        1.253209e-17   -6.654689e+00          2.562490e+00
         Factories      Winnning_Margin              Deaths
     -4.491461e-04       -3.811248e+00        2.551299e-05
>
```

CHOW TEST

We obtain a significant F-statistic 137.82 with a p-value less than 2.2e-16. P-value being less than 2.2e-16 the evidence against the null hypothesis is extremely strong, and we can reject the null hypothesis with a high degree of confidence. Thus, we can conclude that there exists a **Structural Break** in the relationship of Amount of Silicon Dioxide along with other regressors which are SDP, Gini, Politicalparties_seat, Factories, Winning Margin and Deaths.

```
> sctest(params, type = "Chow")

        Chow test

data:  params
F = 137.82, p-value < 2.2e-16
```

T-TEST | HYPOTHESIS TESTING

When we run a t-test, we check the p-values for each predictor variable which is depicted by $Pr(>|t|)$ in the summary of the regression model that we run. When our p-value comes out to be less than 0.05, we reject our null hypothesis and say that the predictor variable we are running the test on is statistically significant. In our case, we obtain all of the values as less than 0.05 with all the newly chosen variables in order to enhance the base model shown in the initial slides having it less than 2e-16 which gives us a strong confidence to reject the null Hypothesis.

| Coefficients	$Pr(>|t|)$
sdp	8.260e-06
Gini	-6.655e+00
Politicalparties_seat	<2e-16
Factories	<2e-16
Winning_Margin	<2e-16
Deaths	<2e-16

MONTE CARLO TEST

Thus, from the regression, we can conclude that there is a statistically significant negative relationship between the predictor variable (SDP) and the outcome variable (Amount.of.Silicon.dioxide). Specifically, the regression coefficient for SDP is estimated to be **-6.826e-06**, which means that for every unit increase in SDP, the amount of silicon dioxide decreases by this amount on average.

Further, with the help of results from the Monte Carlo Test, one can say that the estimated coefficients are reliable and not heavily influenced by the specific characteristics of the original dataset. Therefore, giving us some confidence in the validity of the results.

Finally, looking at the histogram, one can say that estimated coefficients are symmetrically distributed around a central value. Next, we may infer that the estimated coefficients are relatively precise and consistent due to the narrow cluster of the histogram.

```
====================================================
Statistic   N     Mean    St. Dev.   Min     Max
----------------------------------------------------
beta_1     500  -0.0001   0.013    -0.046  0.043
beta_0     500  12.379    0.087    12.111  12.659
----------------------------------------------------
```


Histogram of estimates_DT[, beta_1]
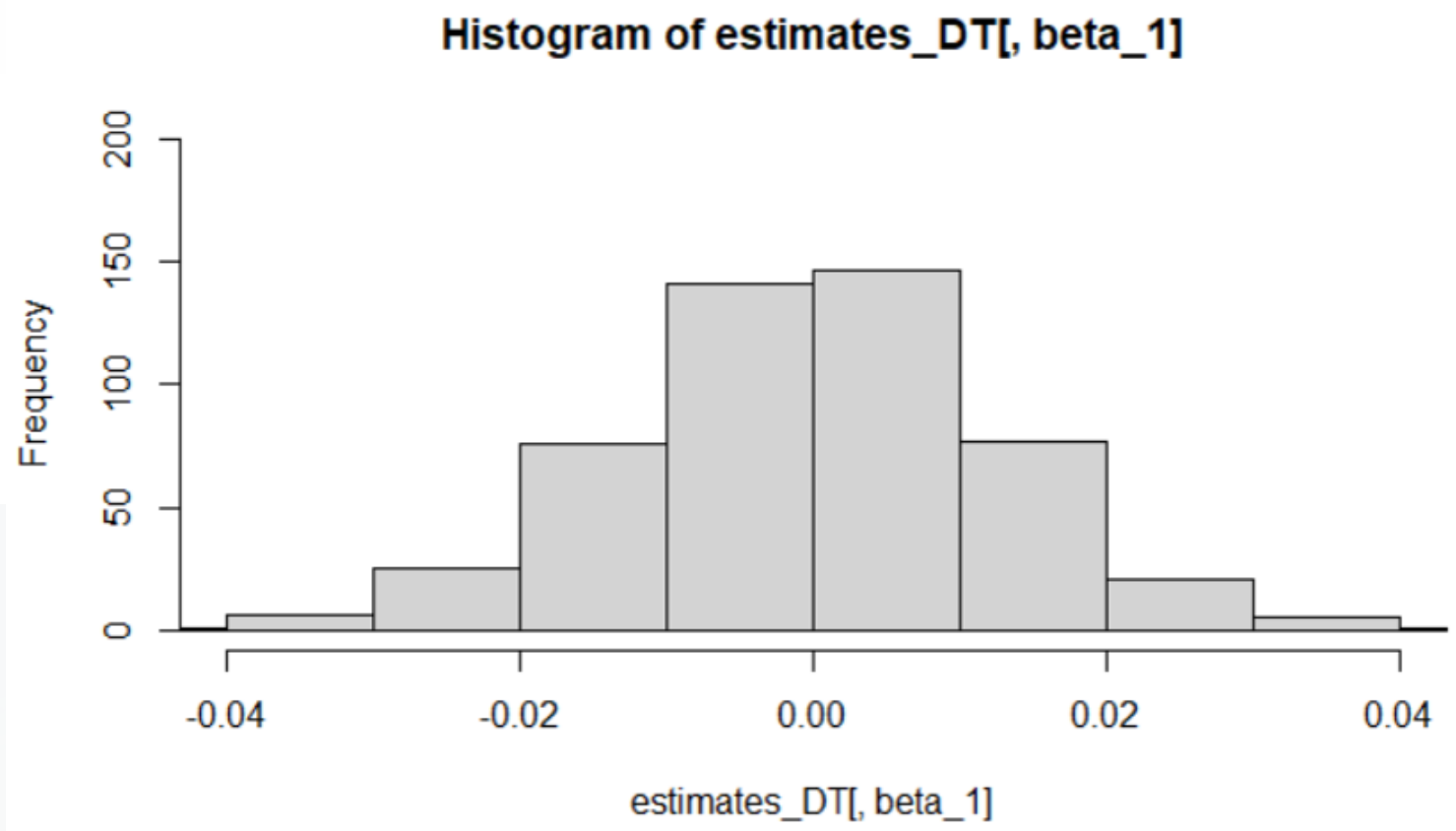
Enhancing the Model: Adding New Variables to Study

- The number of political parties is a representation of how many different opinion groups are being formed per seat. If this number is more, it is going to result in a more diversely opinionated background with higher representation of different voices

- The winning margin represents the political inequality in society as it shows how powerful a single candidate is. If the number is high and implying that candidates are winning with a higher margin, it could imply that people are not having enough choices or that differences of opinions between different groups are not accommodated. It could also imply a situation where the candidate is using unfair means to win the election.

- The Factory variable is representing the industrial growth in a state. The higher the number of factories, the more SiO_2 will be pumped into the air and would contribute more towards air pollution. One might expect the SiO_2 concentration to go down in ground water and up in air quality based on the data.

- Deaths being contributive to the concentration of SiO_2 could depend on how the body is being disposed of. Burials could be the leading causes in increasing the SiO_2 concentration in the ground water

CONCLUSIONS

Winning_Margin

Thus, we can conclude that Winning_Margin plays a significant role in deciding the Amount of Silicon Dioxide in water. The negative coefficient probably depicts that as the margin of winning increases the government tends to put lesser efforts in controlling the amount of silicon dioxide in ground water

Politicalparties_seat

Political parties per unit seat play a very notable role in deciding the Amount of Silicon Dioxide. This might be due to the fact that wherever there are more political parties per seat, there are more voices of people that can be raised due to which there might be an efficient water management system present there thus decreasing the use of groundwater. Further, according to the study conducted by Arina Khan (paper linked at end), it can be said that due to nonuse of ground water and its stagnation causes it to absorb silica from rocks and soil and thus get enriched in it.
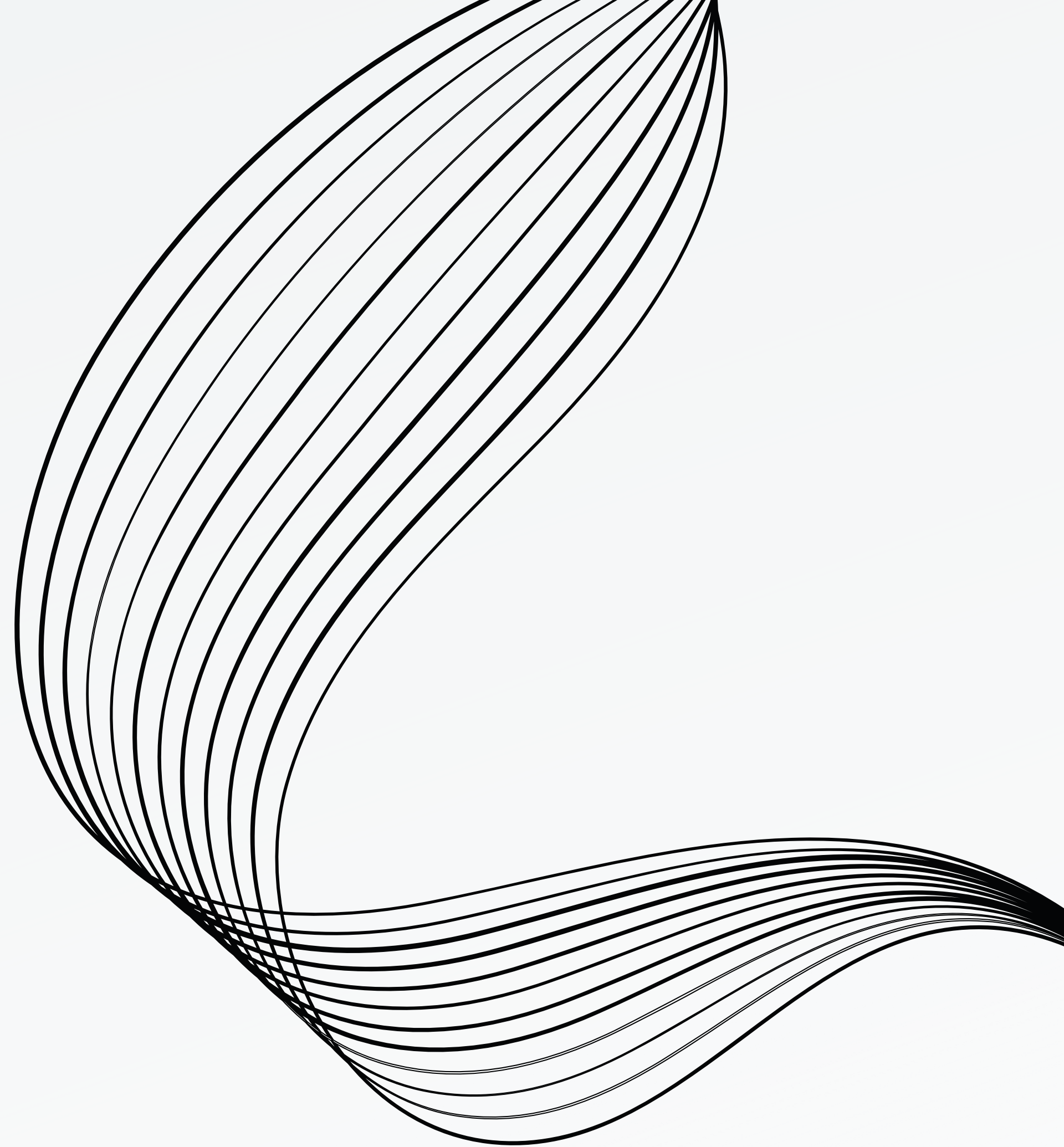
Factories

Silicon Dioxide (Silica) is an aerosol that is mostly released as dust in the air. Further, Silica being non soluble in water and forming Hydrogen Silicate when present in water under optimum temperatures(>29*C). It can be said that probably the factories tend to reduce it in ground water by utiling the ground water in their processes and not releasing silica enriched water back.

Deaths

Silica is present in the Human body in very low amounts after the body is dumped which might get in contact with groundwater and thus increase the value of the Amount of Silicon Dioxide which can be confirmed by the negligible coefficient of the Death regressor in our model

THANK YOU!

REFERENCES

- https://crsorgi.gov.in/annual-report.html
- https://link.springer.com/article/10.1007/s13201-014-0164-z
- https://eci.gov.in/statistical-report/statistical-reports/
- https://www.rbi.org.in/Scripts/PublicationsView.aspx?id=17697