GROUND WATER QUALITY

GROUP 13

SIDDHARTH RAJPUT

DHRUV SOOD

SARVAJEETH U K

ISHANI LAHA

INTRODUCTION

AIM

- In this project, we are trying to test the Environmental Kuznets Curve (EKC) Hypothesis for Ground Water Quality in India
- The Environmental Kuznets Curve (EKC) is an economic theory that suggests that as a country's income increases, its environmental impact will initially increase, but then eventually decrease

DATASET, POWER INEQUALITY & CONTROL VARIABLES

- The chosen dataset has been taken from a span of 18 consecutive years i.e. 2000 to 2018
- Power variables are used to analyze and measure the distribution of power in a society and can affect EKC in many ways
- In the context of EKC, control variables are used to account for the effects of other factors that may affect the relationship between income and environmental quality

VARIABLE DESCRIPTION & SUMMARY

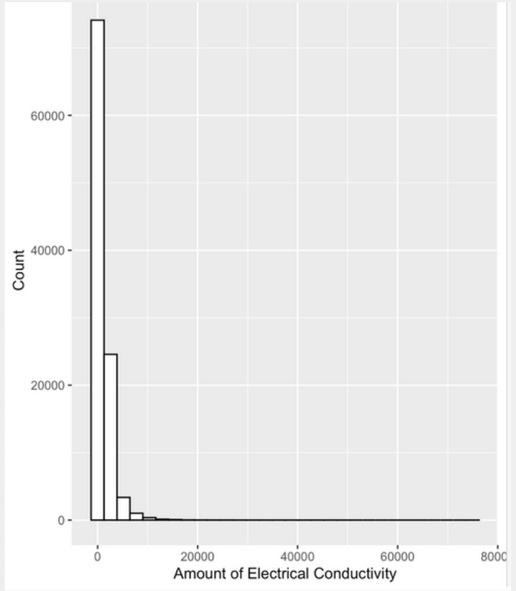
VARIABLE	DESCRIPTION	ACRONYM	TYPE
Electrical Conductivity	Chosen Environment Quality measure, indicates the amount of dissolved ions and minerals in groundwater (S/m)	Amount.of.Electrical.Conductivity	Dependent
Male & Female Passout Metric	Power Variable that reflect gender inequality and access to education, health etc, which are closely associated to avalaibility of clean water resources	ratio	Independent (Power)
Average Voters Turnout	Power Variable that reflect the level of civic engagement of the general public and their awareness towards environmental issues	Turnout_Average	Independent (Power)
Annual Rainfall	Control Variable that could explain impact of rainfall on groundwater quality as it can lead to dilution or enrichment of pollutants (in mm)	ANNUAL.RAINFALL..Millimeters.	Independent (Control)
Telephone Density	Power Variable that could explain impact of fertilisers on groundwater quality as it can leach into the soil and effect its quality (/sq km)	Telephones per sq Km	Independent (Power)
Fertilizer Content	Control Variable that reflects the level of technological development in the area, which translates to better investment in environmental protection(kg/HA)	TOTAL.PER.HA.OF.NCA..Kg.per.ha.	Independent (Control)
SDP	Per Capita value of goods & services produced in a state within a year (in Rupees)	SDP	Independent
Gini Index	The difference in income between the extremes of the economic society. A higher value signifies a greater inequality. (in the range (0,1))	Gini.Index_new	Independent

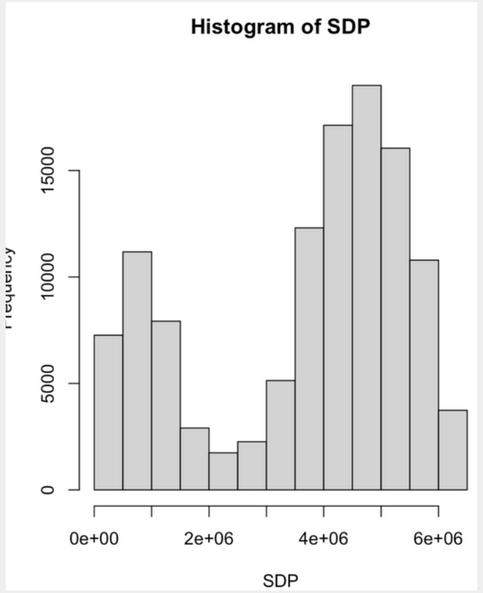
RESULTS FROM DATA ASSIGNMENT I

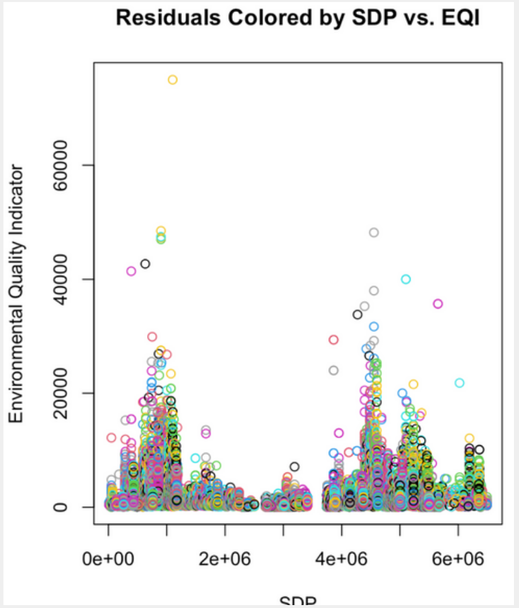
Regression Model

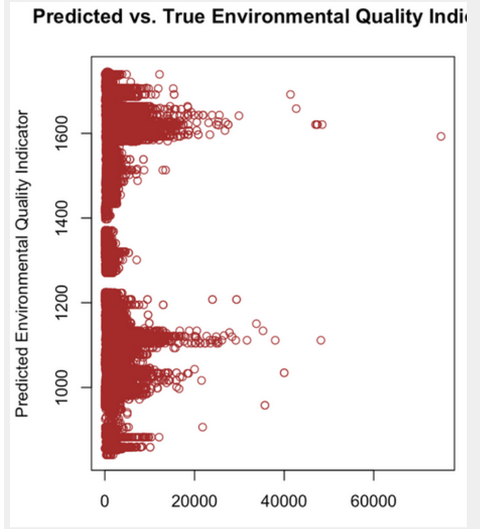
$$t = \alpha_0 + \alpha_1 SDP_{i,t} + \alpha_2 SDP_{i,t}^2 + \alpha_3 SDP_{i,t}^3 + \alpha_4 GINI_i + \gamma_{i,t}$$

Coefficients	Estimates	Std. Error	t value	Pr(>|t|)
(Intercept)	2.500e+03	3.427e+01	72.950	<2e-16
SDP	-2.868e-04	3.368e-05	-8.514	<2e-16
SDP^2	1.621e-11	1.185e-11	1.368	0.171
SDP^3	1.593e-18	1.186e-18	1.343	0.179
Gini Index	-2.187e+03	8.391e+01	-26.061	<2e-16









RESULTS FROM DATA ASSIGNMENT II

Regression Model

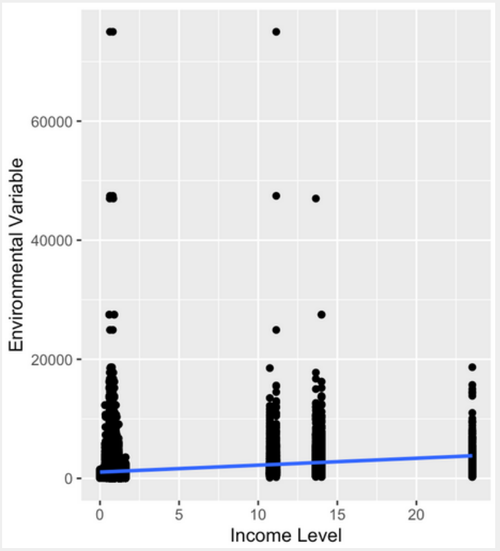
$$t = \alpha_0 + \alpha_1 SDP_{i,t} + \alpha_2 SDP_{i,t}^2 + \alpha_3 SDP_{i,t}^3 + \alpha_4 GINI_i + y_{i,t}$$

Coefficients	Estimates	Std. Error	t value	Pr(>|t|)
(Intercept)	3.278e+03	3.714e+02	8.827	< 2e-16
SDP	-5.859e-04	1.797e-04	-3.260	0.00111
SDP²	-2.167e-10	5.310e-11	-4.081	4.49e-05
SDP³	5.402e-17	4.738e-18	11.401	< 2e-16
Gini Index_New	2.986e+03	4.234e+02	7.052	1.81e-12
TurnoutAverage	1.692e+01	1.313e+00	12.886	< 2e-16
Ratio	1.037e+01	3.396e+00	3.054	0.00226
ANNUAL.RAINFALL	-6.087e-01	1.951e-02	-31.199	< 2e-16
TOTAL.PER.HA.OF.NCA	-2.156e+00	1.322e-01	-16.317	< 2e-16
Telephones per Km2	-2.243e+01	2.503e+00	-8.960	< 2e-16

HYPOTHESIS TESTING I

DEFINITION

- **H0:** The pass metric ratio has no significant effect on the amount of electrical conductivity.
- **HA:** The pass metric ratio has a significant effect on the amount of electrical conductivity.
- Let's assume a **significance level** of **0.05 (5%)**.



TESTING

- The **regression coefficient** of the ratio variable is 0.01975 and the **standard error** is 3.591
- **t-statistic** = (0.01975 - 0) / 3.591 = 0.0055
- **p-value** = 0.9956

INTEPRETATIONS

The p-value (0.9956) is greater than the significance level (0.05), which means we fail to reject the null hypothesis. Therefore, we can conclude that the ratio variable has no significant effect on the amount of electrical conductivity.

CONCLUSION

Based on the hypothesis testing, we can say that the ratio variable is not a significant predictor of the amount of electrical conductivity in this regression model.
The p-value associated with the calculated t-value of 0.0055 is 0.9956, which means there is a 99.56% probability that we would observe a t-value as extreme as 0.0055 or more extreme, assuming the null hypothesis is true.

since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis and conclude that the ratio variable has no significant effect on the amount of electrical conductivity. So we do consider the value in making our decision.

HYPOTHESIS TESTING II

DEFINITION

- **H0:** The Telephones per sq Km variable has no significant effect on the amount of electrical conductivity.
- **HA:** The Telephones per sq Km variable has a significant effect on the amount of electrical conductivity.
- Let's assume a significance level of 0.05 (5%).

TETSING

- The regression coefficient of the ratio variable is 16.79 and the standard error is 3.118
- **t-statistic** = (16.79 - 0) / 3.118 = 5.386
- **p-value** = 7.30e-08

INTEPRETATIONS

Using a t-distribution table or software, we can find the p-value associated with this t-value. The p-value turns out to be 7.30e-08.The p-value (7.30e-08) is much smaller than the significance level (0.05), which means we reject the null hypothesis. Therefore, we can conclude that the Telephones per sq Km variable has a significant effect on the amount of electrical conductivity.Based on the hypothesis testing, we can say that the Telephones per sq Km variable is a significant predictor of the amount of electrical conductivity in this regression model.

CONCLUSION

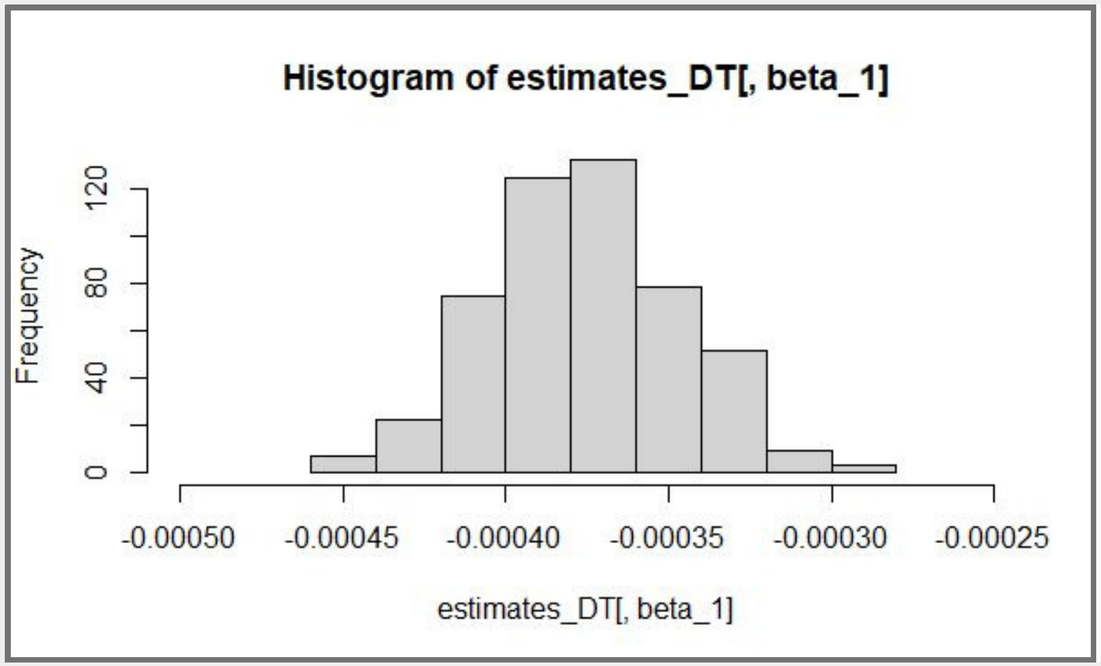
- The p-value associated with the calculated t-value of -5.386 is 7.30e-08, which means there is a very small probability of observing a t-value as extreme as -5.386 or more extreme, assuming the null hypothesis is true.
- In conclusion, since the p-value is smaller than the significance level of 0.05, we reject the null hypothesis and conclude that the Telephones per sq Km variable has a significant effect on the amount of electrical conductivity.

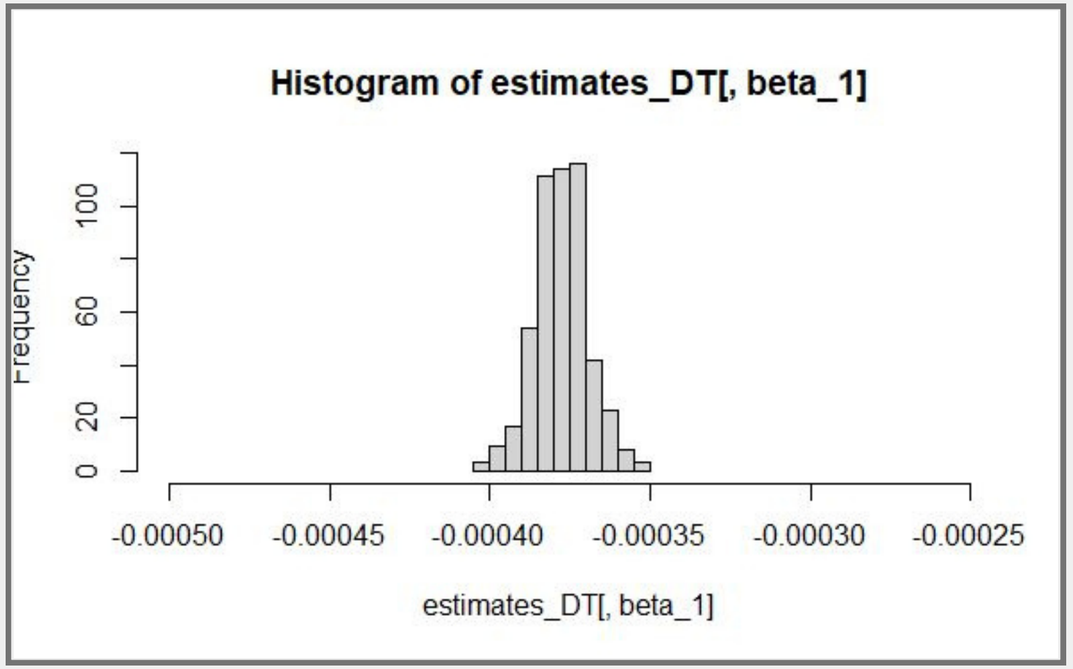
MONTE CARLO SIMULATIONS

True Values

β0= 2399

β1= -0.0003773





Statistic	N	Mean	St. Dv	Min	Max
β0	500	-0.0004	0.00003	-0.0005	-0.0003
β1	500	2390.881	110.392	2060.891	2686.613

Statistic	N	Mean	St. Dv	Min	Max
β0	500	-0.0004	0.00001	-0.0004	-0.0004
β1	500	2401.256	31.595	2290.889	2478.875

CONCLUSION

- After reaching a certain threshold, the impact of SDP and EQI becomes non-linear, and a positive correlation is observed between them. The relationship between the two is not only linear but also follows a complex U-shaped pattern, as indicated by the positive value of SDP2.

- To eliminate the bias in the data, we introduced some control variables such as the use of fertilizers and annual rainfall, and analyzed the impact of various power inequality variables on the chosen Environmental Quality Indicator. By conducting hypothesis testing, we found that some of these variables significantly affect EQI.

- The inclusion of control variables allowed us to obtain a more accurate estimate of the effect of SDP on EQI, indicating that the control variables help remove bias from the data.

- Our analysis showed that SDP has a negative impact on EQI, specifically on the amount of electric conductivity. This means that as income increases, electric conductivity decreases, leading to improved groundwater quality.

- In our enhanced model, we included the state-wise Gini index, which was found to have a higher positive correlation value (4183) than the district-wise Gini index, which had a negative correlation with EQI when other variables were held constant. This difference may be attributed to lower granularity in state-wise data, which leads to such variations.

CONCLUSION

- The standard errors for all the variables were very small, indicating that the coefficient estimates are more precise and reliable.

- Based on observations, it can be concluded that there is a significant relationship between TeleDensity and Electric Conductivity. An increase in TeleDensity leads to improved communication and connectivity, which can have a positive impact on the future Environmental Quality. On the other hand, a negative relationship was found between Voter Turnout Average and Electric Conductivity. This may be due to the fact that people are currently more concerned about demanding better infrastructure and education from the government, and may not be giving as much attention to improving the Environmental conditions.

- We performed Bartlett test for four state groups- south, east, west north to test the equality of variances of the variable amount of electrical conductivity of the four state groups. The p-value associated with the test was extremely small (e-16) indicating the small evidence to suggest the variance of amount of electrical conductivity differ across the state groups

CHOW TEST & T TEST

- We Performed the Chow Test by dividing the data into North,South & East,West as we did in Quiz 3.
- We did the same for the T test also. We have used the t.test() Function to perform the T test

"There is no evidence of structural change in both the sets as the p_value is very Low

North-South

```
Chow test statistic: 450.187
> cat("p-value:", format(p_va
p-value: 0e+00
```

```
data:  Amount.of.Electrical.Conductivity by DNorth
t = -29.2, df = 18004, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -690.9314 -604.0068
sample estimates:
mean in group 0 mean in group 1
       1003.436        1650.906
```

East-West

```
Chow test statistic: 43635.94
> cat("p-value:", format(p_va
p-value: 0e+00
```

```
data:  Amount.of.Electrical.Conductivity by DEast
t = 28.14, df = 25644, p-value < 2.2e-16
alternative hypothesis: true difference in means
95 percent confidence interval:
 469.3640 539.6443
sample estimates:
mean in group 0 mean in group 1
      1395.3269        890.8228
```