Link Between Agricultural Yields and Infant Deaths: A District Level Analysis for India

Akshat Saini (2020019)
Nakul Thureja (2020528)
Mortala Gautam Reddy (2020445)
Ayush Raje Chak (2020502)
Rohan Kulkarni (2020537)

Dependent Variable - FEVER

Our dependent variable is number of infants dying from fever as a percentage of all infants deaths, to explain this phenomenon we tried to find out why an infant may get infected with fever and then applied our domain knowledge of medicine to narrow down the causes to a few which can act as reliable indicators of whether an infant will die from fever instead of just being infected.

Variable Description (Independent)

Variable	Description
gdp	State Wise GDP
tap	District Wise Tap Water Access (Percentage of Households) as of 2019
beds	State Wise Number of Hospital Beds (as of 2020)
v30	Percentage of newborns breastfed within 1 hour (to total live births)
v28	Percentage of newborns having weight less than 2.5 kg (to newborns weighed at birth)
v4	Pregnant women received the second dose of tetanus-toxoid vaccine (TT2) or booster
WQI	State Wise Water Quality Index (added by self) Source: https://cpcb.nic.in/nwmp-data-2019/

Data Summary (for Rabi)

Variable	N	Mean	Median	Mode	SD	Minimum	Maximum
gdp	9259	56682070	45456434	90824131	40121900	725869	180704575
tap	9259	20.96	9.89	0.24	24.38729	0.01	98.98
beds	9259	104313	64939	64939	87739.23	1790	281402
v30	9259	87.22	91.9	97.8	13.62668	0	120.4
v28	9259	14.91	12.8	12	9.97	0.5	90.3
v4	9259	38365	31125	38292	28262.61	249	187255
WQI	9259	3.78	2.4029	2	3.99	0.16	40.5

Data Summary (for Kharif)

Variable	N	Mean	Median	Mode	SD	Minimum	Maximum
gdp	10639	59300143	48623018	90824131	41542556	725869	180704575
tap	10639	21.57	11.77	0.24	24.3	0.01	98.98
beds	10639	108696	83230	64939	87264.39	1790	281402
v30	10639	87.40	92.08	97.8	13.4	0	120.4
v28	10639	15.27	13	12	10.36	0.5	90.3
v4	10639	38271	31410	38384	27809.23	249	187255
WQI	10639	3.5389	2.4029	2	3.8	0.16	40.5

Data Assignment 2 Findings

After completing data assignment 2 , this is the best model that we arrived at, which could explain our dependent variable in a sufficient manner

$$v45_{i,t} = \beta_0 + \beta_1 \cdot log(gdp_{i,t}) + \beta_2 \cdot log(tap_{i,t}) + \beta_3 \cdot log(beds_{i,t}) + \beta_4 \cdot index_{i,t} + \beta_5 \cdot v30_{i,t} + \beta_6 \cdot v28_{i,t} + \beta_7 \cdot log(v4_{i,t}) + \beta_8 \cdot log(WQI_{i,t}) + u_{i,t}$$

Our findings from using this regression model are summarized in the following slides

For Rabi model

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	-0.01	-2.83	0.005	0.005
log(gdp)	-3.36	0.13	<2e-16	0.13
log(tap)	-0.35	-13.16	<2e-16	0.03
log(beds)	1.69	14.77	<2e-16	0.11
v30	0.003	1.09	0.28	0.004
v28	-0.05	-10.08	<2e-16	0.005
log(v4)	-0.12	-1.79	0.07	0.07
log(WQI)	-0.02	-0.33	0.74	0.05

Residual Standard Error: 4.462 on 9250 degrees of freedom
Multiple R-squared: 0.21
Adjusted R-squared: 0.21
F-statistic: 302.3 on 8 and 9250 DF
P-value: < 2.2e -16

For Kharif Model

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	0.02	3.12	0.002	0.01
log(gdp)	-2.89	-23.40	<2e-16	0.12
log(tap)	-0.36	-13.93	<2e-16	0.03
log(beds)	1.35	12.15	<2e-16	0.11
v30	0.007	2.21	0.03	0.003
v28	-0.04	-9.4	<2e-16	0.004
log(v4)	-0.15	-2.33	0.02	0.06
log(WQI)	0.03	0.53	0.6	0.05

Residual Standard Error: 4.396 on 10630 degrees of freedom
Multiple R-squared: 0.19
Adjusted R-squared: 0.19
F-statistic: 315.9 on 8 and 10630 DF
P-value: < 2.2e -16

Explaining the Findings

The trends across both the cropping seasons remain largely the same for every independent variable with the exception of index and log(WQI).

Index has a positive relationship with v45 in the kharif season while a negative relation in rabi which is quite interesting since kharif season also includes the time of a seasonal change, we experience a shift in the weather, which as we know is conducive environment for viruses to thrive and spread easily.

A similar trend is present for log(WQI), this can perhaps be attributed to the increased humidity in kharif season making it easier for viruses to survive long enough to infect people and babies.

Introduction to the Project

In data assignment 2, the states were divided into six different zones based on their locations, and then we tried to test whether there is a structural break across these zones.

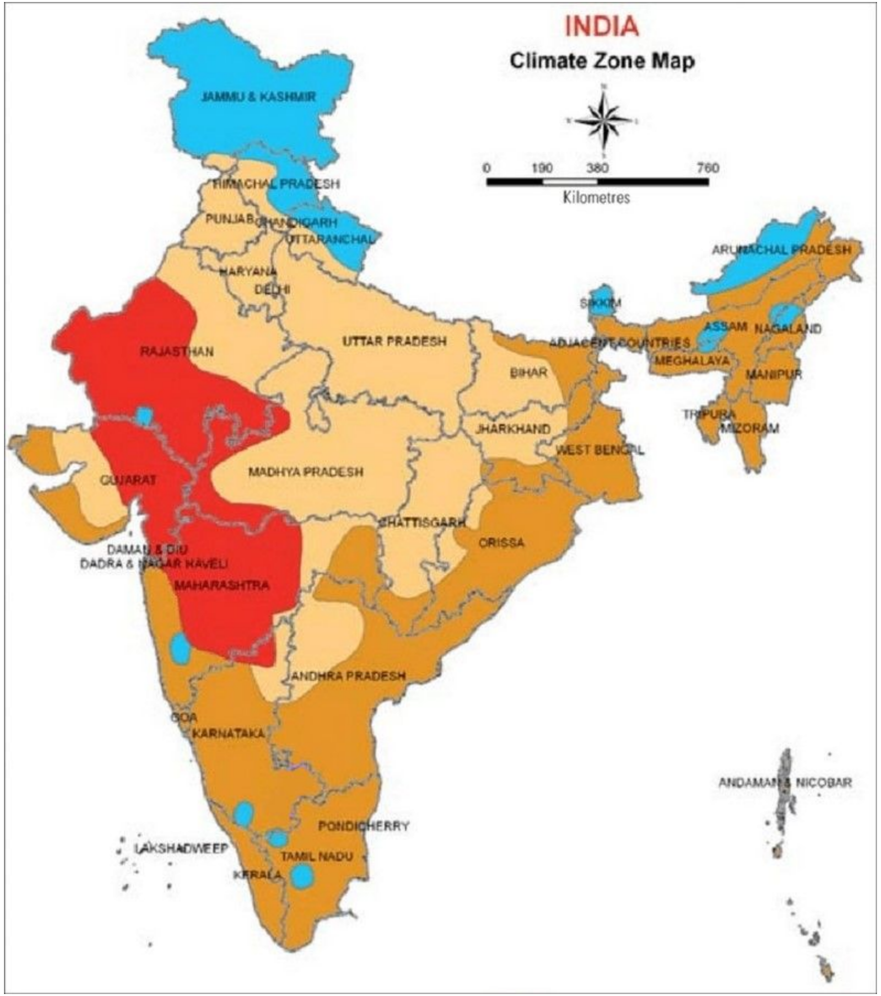
So we built upon this idea along with the suggestions of the model that different kinds of weather and climate might have different effects on the infant deaths due to fever and divided the districts into four different categories based on their climate

- Warm and Humid Climate
- Cold Climate
- Composite Climate
- Hot and Dry Climate

And tried to test whether there is a structural break across these climatic zones.

Distribution of Climatic Zones in India

Climate Type	Key
Warm and Humid climate	
Cold climate	
Composite climate	
Hot and Dry Climate	



Thought Process

India is a diverse country with a wide range of climate zones and we thought that similar to how geographic boundaries might affect the health indicator, we thought that perhaps climate zones might also affect the health indicators in some manner.

As already stated this hypothesis of ours got a boost in preliminary discussions due to the change in the findings for kharif and rabi seasons for some dependent variables.

Modification to test Hypothesis

We first categorized districts into climate zone

$$v45_{i,t} = \beta_0 + \beta_1 \cdot log(gdp_{i,t}) + \beta_2 \cdot log(tap_{i,t}) + \beta_3 \cdot log(beds_{i,t}) + \beta_4 \cdot index_{i,t} + \beta_5 \cdot v30_{i,t} + \beta_6 \cdot v28_{i,t} + \beta_7 \cdot log(v4_{i,t}) + \beta_8 \cdot log(WQI_{i,t}) + \beta_9 \cdot D_{i,cold} + u_{i,t}$$

Here $\boldsymbol{\beta}_9$ is the coefficient associated with the dummy variable $D_{i,cold}$.

Since $\boldsymbol{\beta}_9$ represents the difference between average Health outcomes between cold climatic zones and the country. Therefore it accounts for the structural break in mean outcomes level across different climatic zones.

Dummy Variables

Variable	Description	Acronym
Warm and Humid climate	Has value 1 for districts having Warm and humid climate, else 0	dwh
Cold climate	Has value 1 for districts having cold climate, else 0	dcd
Composite climate	Has value 1 for districts having composite climate, else 0	dcp
Hot and Dry Climate	Has value 1 for districts having cold climate, else 0	dhd

Classification of climates and district-wise data derived from **NBC climatic region(Bureau of Indian Standard, 2016)**

NULL HYPOTHESIS

Similar to what we did in Data Assignment 2 with regards to Geographical zones, we will aim to do the same with climatic zones.

We will use the t-statistic to test the null and find if there is a structural difference between the climatic zones. Here our null hypothesis H_0 is that v45 is not dependent on the climate zones i.e. $\boldsymbol{\beta}_9$

Data Summary

Variable	N	Mean	Median	Mode	SD	Minimum	Maximum
gdp	26734	58401326	48623018	90824131	40682225	725869	180704575
tap	26734	21.70	11.40	0.24	24.94	0.01	98.98
beds	26734	107098	64939	64939	88079.56	1790	281402
v30	26734	87.52	92.20	97.8	13.35	0	120.4
v28	26734	15.0	12.8	10.9	10.05	0.5	90.3
v4	26734	38313	31262	38292	28163.3	249	187255
WQI	26734	3.5537	2.4	2	3.97	0.16	40.5

DWH (Warm and Humid)

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	-0.003	-2.29	0.022	0.002
log(gdp)	-3.18	-43.60	<2e-16	0.07
log(tap)	-0.40	-25.60	<2e-16	0.02
log(beds)	1.44	22.82	<2e-16	0.06
v30	-0.01	-2.86	0.03	0.002
v28	-0.06	-20.32	0.004	0.003
log(v4)	-0.23	-5.78	2e-16	0.04
log(WQI)	-0.04	-1.37	7.63e-09	0.03
dwh	-1.36	-23.33	2e-16	0.06

Residual Standard Error: 4.282 on 26385 degrees of freedom
Multiple R-squared: 0.22
Adjusted R-squared: 0.22
F-statistic: 804.2 on 9 and 26385 DF
P-value: < 2.2e -16

The confidence value for dwh is less than 0.05, so we can reject the null hypothesis

DCD (Cold)

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	-0.005	-3.03	0.002	0.002
log(gdp)	-3.06	-41.44	<2e-16	0.074
log(tap)	-0.34	-21.66	<2e-16	0.016
log(beds)	1.43	22.34	<2e-16	0.06
v30	0.004	2.14	0.03	0.002
v28	-0.04	-16.34	<2e-16	0.003
log(v4)	-0.13	-3.33	0.001	0.04
log(WQI)	-0.01	-0.33	0.74	0.03
dcd	-0.04	-0.12	0.91	0.31

Residual Standard Error: 4.326 on 26385 degrees of freedom
Multiple R-squared: 0.20
Adjusted R-squared: 0.20
F-statistic: 728.7 on 9 and 26385 DF
P-value: < 2.2e -16

The confidence value for dcd is greater than 0.05, so we can accept the null hypothesis

DCP (Composite)

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	-0.005	-2.90	0.004	0.002
log(gdp)	-2.92	-39.48	<2e-16	0.07
log(tap)	-0.33	-21.44	<2e-16	0.02
log(beds)	1.29	20.03	<2e-16	0.06
v30	0.002	0.81	0.42	0.002
v28	-0.05	-16.63	<2e-16	0.003
log(v4)	-0.18	-4.54	5.72e-06	0.04
log(WQI)	-0.05	-1.69	0.09	0.03
dcp	0.78	13.60	<2e-16	0.06

Residual Standard Error: 4.311 on 26385 degrees of freedom
Multiple R-squared: 0.21
Adjusted R-squared: 0.21
F-statistic: 754.3 on 9 and 26385 DF
P-value: < 2.2e -16

The confidence value for dcp is less than 0.05, so we can reject the null hypothesis

DHD (Hot and Dry)

Variable	Estimator (Slope)	T value	P value	Standard Error
Index	-0.004	-2.70	0.007	0.002
log(gdp)	-3.31	-43.39	<2e-16	0.08
log(tap)	-0.38	-24.06	<2e-16	0.02
log(beds)	1.61	24.61	<2e-16	0.07
v30	0.0003	0.14	0.89	0.002
v28	-0.05	-18.48	<2e-16	0.003
log(v4)	-0.15	-3.86	0.0001	0.04
log(WQI)	0.02	0.56	0.55	0.03
dhd	0.94	12.20	< 2e-16	0.08

Residual Standard Error: 4.314 on 26385 degrees of freedom
Multiple R-squared: 0.20
Adjusted R-squared: 0.20
F-statistic: 749.3 on 9 and 26385 DF
P-value: < 2.2e -16

The confidence value for dhd is less than 0.05, so we can reject the null hypothesis

Findings

After running the t Test to check our hypothesis, we found some surprising results, the only climate group that did not differ from the mean is the colder climate zone indicating it had no structural break from the mean

For the other 3 climate dummy variables we reject their null hypothesis due to the aforementioned reasons which indicates that their beta-ols is not zero and hence have a structural break with the mean which means that we can successfully conclude that climate also plays a role in influencing our assigned health indicator.

Thank You

Group 7:
Akshat Saini: 2020019
Nakul Thureja: 2020528
Mortala Gautam Reddy: 2020445
Ayush Raje Chak: 2020502
Rohan Kulkarni: 2020537