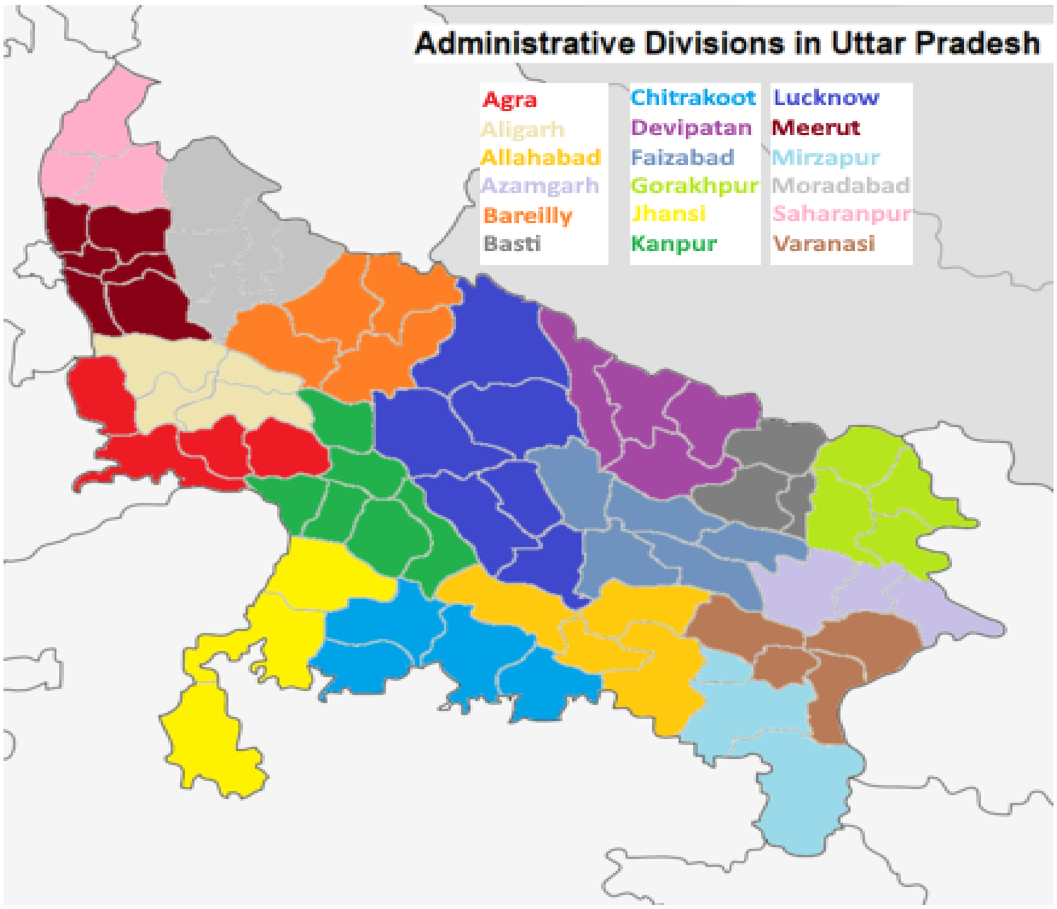
STATISTICAL METHODS - 3

ANALYSIS OF FEASIBILITY OF THE SUBDIVISION OF UTTAR PRADESH



INTRODUCTION

In this project, we cast a statistical eye on the proposal of former chief minister of Uttar Pradesh Kumari Mayawati to divide the state into four separate states, based on geographical, linguistic and cultural similarities for the purpose of better administration of the state. First we look at the relationship between the sizes of various states and their prosperity. So, in the first part of the project we do a multiple linear regression hypothesis testing using python where our DEPENDENT variable is the GSDP per capita of each state and there are three INDEPENDENT variables: Number of districts in the state, average size of a district in the state and the average population of a district in the state. In the data set we have entries for all the 28 states and 8 Union Territories except the smaller UTs of - Daman and Diu, Dadra and Nagar Haveli and Lakshadweep due to unavailability of up to date data and they would anyways distort the data

In the second part of the project, we perform a MANOVA over the chief development indicators of the various parts. The parts in this MANOVA are similar to but different from Mayawati's proposal. This division was considered because the data published by the UP Government, which is our source in the second part of the project categorised the districts in this manner. (Link to data - dist_dev_indicators_2023.pdf). Here is an image of the division considered. Western, Eastern, Central parts and the Bundelkhand region in the south.

THE MULTIPLE REGRESSION

DATA

State	GSDP per capita	Num of Districts	Area	Population
Andhra Pradesł	2,45,582	26	1,62,968	5,34,02,000
Arunachal Prad	2,25,767	26	83,743	15,80,000
Assam	1,34,590	35	78,438	3,61,59,000
Bihar	59,243	38	94,163	12,92,05,000
Chhattisgarh	1,54,609	33	1,35,191	3,06,38,000
Goa	5,67,346	2	3,702	15,85,000
Gujarat	3,09,691	33	1,96,024	7,26,53,000
Haryana	3,29,778	22	44,212	3,06,94,000
Himachal Prade	2,57,251	12	55,673	75,18,000
Jharkhand	1,06,309	24	79,714	4,01,29,000
Karnataka	3,36,221	31	1,91,791	6,82,56,000
Kerala	2,86,595	14	38,863	3,59,67,000
Madhya Prades	1,44,688	55	3,08,245	8,79,54,000
Maharashtra	2,89,405	36	3,07,713	12,76,84,000
Manipur	1,25,405	16	22,327	32,60,000
Meghalaya	1,38,379	12	22,429	33,87,000

THE CODE

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.stats.diagnostic import het_breuschpagan

df = pd.read_excel("Large State Testing.xlsx")

# Display the first few rows of the dataset to understand its structure
df.head()
```

```python
# Recalculate the independent variables
df['Avg_size_district_km2'] = df['Area'] / df['Districts']
df['Avg_population_district'] = df['Population'] / df['Districts']

# Redefine the dependent and independent variables
X_corrected = df[['Districts', 'Avg_size_district_km2', 'Avg_population_district']]
y_corrected = df['GSDP']

# Add a constant to the independent variables for the regression
X_corrected = sm.add_constant(X_corrected)

# Fit the corrected multiple linear regression model
model_corrected = sm.OLS(y_corrected, X_corrected).fit()

# Display the corrected regression summary
corrected_summary = model_corrected.summary()
print(corrected_summary)
```

BREUSCH-PAGAN TEST - Under the classical assumptions, ordinary least squares is the best linear unbiased estimator (BLUE), i.e., it is unbiased and efficient. It remains unbiased under heteroskedasticity, but efficiency is lost. Before deciding upon an estimation method, one may conduct the Breusch–Pagan test to examine the presence of heteroskedasticity. We do the test for our data in the following manner and find that the data is homskedastic (as the p- value is: 0.15 > 0.05), validating our regression.

```python
# Perform the Breusch-Pagan test for homoscedasticity
bp_test = het_breuschpagan(model_corrected.resid, X_corrected)

# Extract results
bp_stat = bp_test[0]  # Lagrange multiplier statistic
bp_pvalue = bp_test[1]  # p-value for the test

print(bp_stat, bp_pvalue)
```

BP Test results -

```
5.240809778075542 0.1549895957715503З
```

REGRESSION RESULTS -

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.303e+05	4.22e+04	7.818	0.000	2.44e+05	4.17e+05
Districts	-4669.2586	1681.692	-2.777	0.010	-8108.705	-1229.812
Avg_size_district_km2	3.7261	12.814	0.291	0.773	-22.482	29.934
Avg_population_district	0.0139	0.023	0.598	0.555	-0.034	0.061

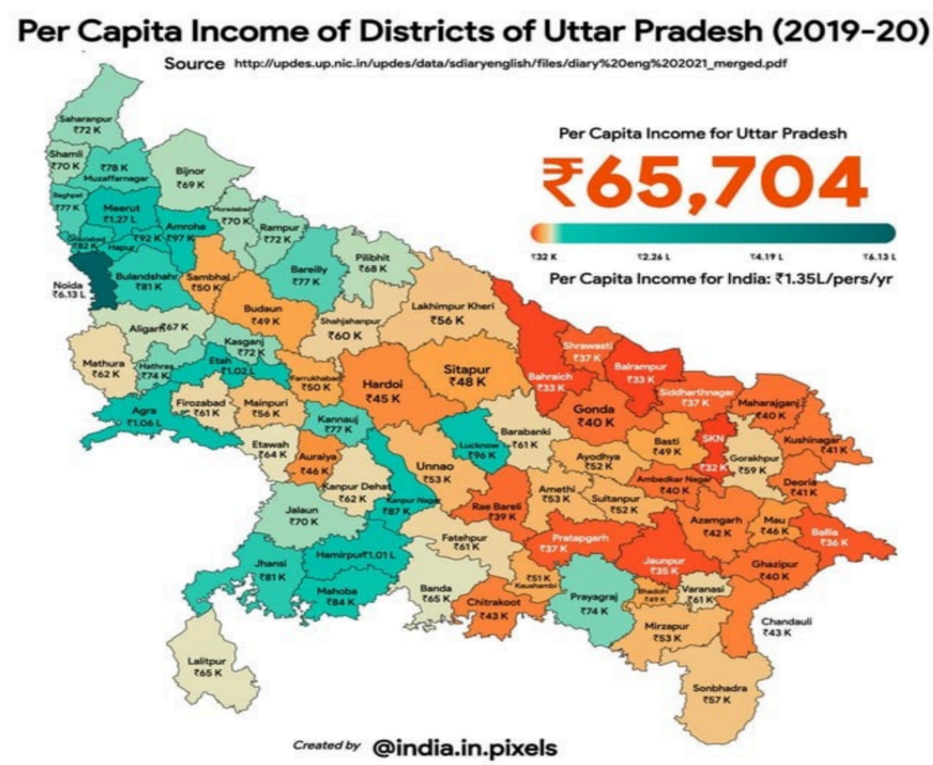
We can derive the following conclusions from the results of the test -

1. **F-statistic**: 3.264 with a p-value of 0.0355, indicating the overall model is statistically significant at the 5% level. Thus not all the regression coefficients are 0.
2. **Constant (Intercept)**: 330,300 (p-value < 0.001), highly significant.
3. **Number of Districts**: -4,669 (p-value = 0.010), statistically significant. This indicates that an increase in the number of districts is associated with a decrease in GSDP per capita.
4. **Average Size of District (km²)**: 3.73 (p-value = 0.773), not statistically significant.
5. **Average Population of District**: 0.0139 (p-value = 0.555), not statistically significant

This gives us evidence in favour of Mayawati's proposal. That the division would be favourable. But on a realistic and practical level there is an additional key problem that we would face that in my opinion should delay this division for the time being.

The income distribution in Uttar Pradesh (UP) exhibits a distinct east-west divide. The western region, with better industrialization, urbanization, and agricultural productivity, has higher incomes. In contrast, the eastern region, characterized by limited infrastructure, low industrial presence, and agrarian challenges, faces lower income levels, reflecting socio-economic disparities across the state.

This is similar to the current infrastructural and technological divide between East and West Germany. West part being better in that case as well.



The east-west income disparity in Uttar Pradesh has historical roots in colonial and post-independence economic policies, geographical factors, and social structures:

1. **Colonial Policies**: During British rule, western UP benefitted from canal-based irrigation systems (e.g., Ganga Canal), enabling higher agricultural productivity. Eastern UP, dependent on monsoon rains, lagged behind.
2. **Green Revolution**: Post-independence, the Green Revolution in the 1960s favored regions with better irrigation and fertile soil, such as western UP, further boosting agricultural income and economic growth there. Eastern UP missed out due to poor irrigation and fragmented landholdings.
3. **Industrial Development**: Industrial hubs like Noida, Ghaziabad, and Agra developed in western UP due to proximity to Delhi and better connectivity. Eastern UP remained agrarian, with few industries

In order to further demonstrate how the division of UP would be unjust towards the eastern part we turn to the second part of the project. And we use the statistical technique of MANOVA.

THE SUBPARTS ARE UNEQUAL

The source of our data are the annual district development indicators published by the Government of Uttar Pradesh available at the following link - updes.up.nic.in/updes/data/dist_dev_indicator/dist_dev_indicators_2023.pdf. We handpicked some of the 7 of the more than 100 indicators that cover different aspects.

They are - Population Density, Number of allopathic hospitals (per lakh), Literacy rate, Number of Higher Senior Secondary School (per lakh) of population, (Per capita) electricity consumption, Number of BSNL mobile connections (per lakh), (Per capita) foodgrain production, (Per capita) value of industrial production, (Per capita) deposits, Percentage of total workers and (Per capita) Net Domestic Product.

MANOVA - Multivariate Analysis of Variance (MANOVA) is a statistical technique used to assess whether there are significant differences in multiple dependent variables across groups defined by one or more categorical independent variables. It extends the ANOVA framework by analyzing several dependent variables simultaneously while accounting for their interrelationships. MANOVA tests the hypothesis that the mean vectors of the dependent variables are equal across groups. It is particularly useful when the dependent variables are correlated, as it considers their combined variance.

DATA

	Population Density	No. of allopathic hospitals per lakh	Literacy rate	No. of schools per lakh (HSS	Per capita electricity consumption	No. of mobile connections per lak	Per capita foodgrains produce
Western Part	930	2.19	67.47	12.11	620.41	4560	260.08
Central Part	785	2.16	68.31	10.82	463.27	2091	251.16
Bundelkhand	329	2.98	69.26	9.89	401.25	1371	482.34
Eastern Part	931	2.78	67.4	12.63	320.55	1807	253.56

Per capita foodgrains produced	Per capita value of production	per capita deposits	percentage of total workers	Per capita NDP(current rices)
260.08	45976.03	70507.57	31.31	93930.5
251.16	20173.88	95825.13	34.27	69337.74
482.34	11887.26	36496.59	39.54	76423.89
253.56	5581.65	39327.84	33.06	49077.76

While the MANOVA did not produce interpretable results due to perhaps the large number of columns in the data and insufficient knowledge on my part about the technique.

Here are the code and results of the MANOVA.

```
import pandas as pd
from statsmodels.multivariate.manova import MANOVA

# Load the data from the first sheet
data = pd.read_excel("Real MANOVA.xlsx", sheet_name=0)

# Rename the first column to "Region" if it's unnamed
data = data.rename(columns={"Unnamed: 0": "Region"})

# Sanitize column names to make them formula-friendly
data.columns = data.columns.str.replace(r'\W+', '_', regex=True)

# Check the cleaned column names
print("Sanitized Column Names:", data.columns)

# Separate dependent and independent variables
independent_var = data["Region"]
dependent_vars = data.drop(columns=["Region"])

# Construct the formula for MANOVA
formula = ' + '.join(dependent_vars.columns) + ' ~ Region'

# Fit the MANOVA model
manova = MANOVA.from_formula(formula, data=data)
manova_results = manova.mv_test()

# Display the results
print(manova_results)
```

Multivariate linear model

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0000	1.0000	0.0000	0.0000	nan
Pillai's trace	2.0009	1.0000	0.0000	-0.0000	nan
Hotelling-Lawley trace	9007199254740992.0000	1.0000	0.0000	0.0000	nan
Roy's greatest root	9007199254740991.0000	1.0000	0.0000	0.0000	nan

Region	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0000	9.0000	-4.7169	-392943688325946112.0000	nan
Pillai's trace	5.7428	9.0000	0.0000	-0.0000	nan
Hotelling-Lawley trace	1747291852842769.7500	9.0000	-15.0000	-970717696023761.0000	nan
Roy's greatest root	1286742750677283.5000	3.0000	0.0000	0.0000	nan

While the statistical method, didn't produce the desired result, by eyeball we can see that even though the social indicators of the different parts are comparable, the economic indicators of the western part, aided in part by the proximity to New Delhi are totally incomparable to the other 3 parts.

CONCLUSION

In much the same way as at the national level, the tax revenues from the southern states help in the development of the northern states that lag behind, comparatively, eastern UP needs the access of revenues of western UP so that the basic industrial setup is achieved in Eastern UP. Therefore, the plan of division of the state into four parts although required at some point as shown by our multiple regression study needs to be put in the backseat at the present moment.

BIBLIOGRAPHY

1. Ramachandra Guha: How India will benefit if Uttar Pradesh is divided into smaller states. https://scroll.in/article/1008419/ramachandra-guha-how-india-will-benefit-if-uttar-pradesh-is-divided-into-smaller-states

2. The case against dividing Uttar Pradesh into smaller states. Shivam Vij. https://theprint.in/opinion/the-case-against-dividing-uttar-pradesh-into-smaller-states/459618/

3. The Division of Uttar Pradesh Into 4 States: A Forgotten Issue. Yatharth Mishra. https://www.thequint.com/amp/story/voices/blogs/uttar-pradesh-divide-four-parts-bsp-sp-elections-mayawati-akhilesh-yadav#read-more

4. UP should learn from Uttarakhand that splits are a mixed blessing. Santosh Kumar. https://m.economictimes.com/opinion/et-commentary/up-should-learn-from-uttarakhand-that-splits-are-a-mixed-blessing/articleshow/10965052.cms

5. Should UP be divided into 4 states for better governance? Analysis by Shashank Pandey. https://youtu.be/PwenaSTPteU?si=1TXipifnNnVlUk6V