Difference-in-Difference (DiD) Regression Model

Final Dataset Structure:

- Outcome variable (Overdoses per capita, Opioid shipments).
- Time variable (Year).
- Policy implementation indicator (policy_state) => 1 if the county is in a state with the policy change and 0 otherwise.
- A time variable that marks the periods before and after the policy change (post_policy) => 1 for years after the policy change and 0 for years before.

Model Definition:

$$Y_{c,t}=\alpha+\psi_c+\beta_1 \cdot post_t+\beta_2 \cdot time_t+\beta_3 \cdot time_t \cdot policy\ state_c +\beta_4 \cdot post_t+\beta_5 \cdot post_t \cdot policy\ state_c+\beta_6 \cdot time_t \cdot post_t+\beta_7 \cdot time_t \cdot post_t \cdot policy\ state_c+\epsilon_{c,t}$$

Proposed Python Script for Model:

```python
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Final dataset columns:
# 'outcome' (Yc,t), 'post_policy' (postt), 'policy_state' (policy_statec),
# 'time' (time_t), and 'county' (ψc)

# Create necessary interaction terms
final_df["time_policy"] = final_df["time"] * final_df["policy_state"]
final_df["post_time"] = final_df["post_policy"] * final_df["time"]
final_df["post_time_policy"] = (
    final_df["post_policy"] * final_df["time"] * final_df["policy_state"]
)

# Formulate regression formula according to our final dataset structure
formula = (
    "outcome ~ post_policy + time + time_policy + post_time + "
    "post_time_policy + C(county)"
)

# Run the regression with county fixed effects
model = smf.ols(formula, data=final_df).fit(
    cov_type="HC1"
)  # Using HC1 for robust standard errors

# Print the results
print(model.summary())
```

Interpretation of Coefficients:

- α: *Baseline level for control counties before the policy change.*
- $\alpha+\beta_1$: *Baseline level for treated counties before the policy change.*
- β_1: *Difference in baseline levels between treated and control counties at the time of the policy change.*
- $\alpha+\beta_4$: *Baseline level for control counties after the policy change.*
- $\alpha+\beta_1+\beta_4+\beta_5$: *Baseline level for treated counties after the policy change.*
- $\beta_1+\beta_5$: *Difference in baseline levels between treated and control counties after the policy change.*
- β_2: *Trend for control counties before the policy change.*
- $\beta_2+\beta_3$: *Trend for treated counties before the policy change.*
- β_3: *Difference in trends between treated and control counties before the policy change.*
- $\beta_2+\beta_6$: *Trend for control counties after the policy change.*
- $\beta_2+\beta_3+\beta_6+\beta_7$: *Trend for treated counties after the policy change.*
- $\beta_3+\beta_7$: *Difference in trends between treated and control counties after the policy change.*

The DiD Estimates:

- β_5: *This is the DiD estimate for the change in the intercept. It measures the immediate level change in the outcome due to the policy, comparing before and after the policy within the treated counties relative to the control counties.*
- β_7: *This is the DiD estimate for the change in the slope. It measures the change in the trend of the outcome due to the policy, comparing before and after the policy within the treated counties relative to the control counties.*