EVALUATING VOTER TURNOUT TRENDS USING SAS MULTIVARIATE ANALYSIS

1. Overview

Client:

A nonpartisan U.S. policy research institute supporting civic engagement and electoral reform

Objective:

To identify and quantify the factors that influence voter turnout across U.S. states using SAS. The goal was to inform future voter outreach strategies and support academic publications on electoral behavior.

2. Background

Voter turnout in the U.S. varies significantly by region, age, race, and income. While descriptive data are abundant, multivariate statistical modeling is required to uncover the relative weight of each factor while controlling for others. SAS was chosen for its strength in handling large survey datasets and its proven use in social science research.

3. Data Summary

Dataset:

National election study dataset from 2016, 2018, and 2020 (N = 45,000+ respondents)

Variables Used:

Variable	Type	Description
Voted (0/1)	Binary	Whether the respondent voted in the most recent election
Age	Continuous	Age in years
Education_Level	Categorical	High school, college, postgraduate
Race_Ethnicity	Categorical	White, Black, Hispanic, Other
Household_Income	Ordinal	Low, Medium, High (quantile-based)
Party_Affiliation	Categorical	Democrat, Republican, Independent
State	Categorical	U.S. state of residence

Registration_Status	Binary	Whether respondent is registered to vote

4. Methodology

Software Used:

SAS 9.4

SAS Procedures Employed:

1. **Data Cleaning & Formatting:**

   o PROC IMPORT, DATA steps for label encoding and filtering

   o Missing value imputation using mean/mode for select fields

   o Dummy variables created for categorical predictors

2. **Descriptive Statistics:**

   o PROC FREQ, PROC MEANS, and PROC UNIVARIATE for distributions

   o Cross-tabulation of voter turnout by key demographic variables

3. **Multivariate Analysis:**

   o PROC LOGISTIC to model likelihood of voting (Voted = 1)

   o Backward elimination used for variable selection

   o Included interaction terms (e.g., Race * Income)

   o Goodness-of-fit assessed via Hosmer–Lemeshow test

4. **Model Validation & Diagnostics:**

   o ROC curve, AUC calculation

   o Decile-based lift analysis

   o Residual diagnostics using INFLUENCE and LEVERAGE

5. Key Results

Predictor	Odds Ratio	p-value	Interpretation
Age	1.02	<0.001	Older respondents more likely to vote

Education_Level (Postgrad)	1.85	<0.001	Higher education linked to increased turnout
Race_Ethnicity (Black)	1.34	0.009	Black respondents had higher turnout controlling for factors
Registration_Status	6.21	<0.001	Registration is the strongest predictor of turnout
Party_Affiliation (Democrat)	1.41	0.015	Democrats more likely to vote than Independents

Model Performance:

- AUC = 0.82

- Hosmer–Lemeshow: $p = 0.69$

- C-statistic showed excellent discrimination

6. Visual Outputs (from SAS)

- ROC Curve and AUC summary

- Predicted turnout probability by age and education

- Interaction plot (Income × Race)

- Decile lift chart for top predictors

7. Deliverables

- SAS .sas program and dataset with model and diagnostics

- Final report (19 pages) including:

    o Research question and variable definitions

    o Model-building process and diagnostics

    o Graphs, tables, and practical insights

    o Interpretation tailored for policy and outreach audiences

- Executive summary (2 pages) with:

    o Policy implications (e.g., focusing on registration drives)

    o Recommendations for outreach segmentation and timing

8. Application & Outcome

- Cited in the client's 2022 whitepaper on civic engagement strategy

- Used by state-level nonprofits to prioritize ZIP codes for voter registration drives

- Informed academic publications submitted to journals in political science and public affairs

9. Strategic Value Delivered

- Provided **quantitative backing for voter outreach priorities**

- Delivered **interpretability and rigor suitable for publication and advocacy**

- Strengthened the client's position as a **data-driven thought leader in electoral policy**