Causal Inference Analysis: ATE Analysis for Smoking Cessation & Legislator Response Patterns

Bibin Joseph

Causal Inference Analysis: Smoking Cessation & Legislator Response Patterns

Overview

This project applies causal inference techniques to two datasets:

1. **National Health and Nutrition Examination Follow-up Study (NHEFS)** – to estimate the effect of smoking cessation on weight change and blood pressure.
2. **Legislator Email Response Dataset** – to examine whether the race of a legislator influences their likelihood of responding to constituent emails.

The analyses focus on estimating the **Average Treatment Effect (ATE)** using unadjusted methods and summarizing key findings.

The Effect of Smoking Cessation on Health

The dataset is first cleaned by selecting relevant variables, transforming categorical variables into factors, and labeling them for clarity.

Descriptive Statistics

To understand differences between those who quit smoking and those who did not, we summarize the baseline characteristics.

Table 1: Summary Statistics Stratified by Smoking Cessation

	level	No	Yes
n		1201	428
Sex (%)	Male	562 (46.8)	237 (55.4)
	Female	639 (53.2)	191 (44.6)
Age in 1971 (mean (SD))		42.92	46.70
		(11.89)	(12.52)
Race in 1971 (%)	White	1024 (85.3)	390 (91.1)
	Black or Other	177 (14.7)	38 (8.9)
Total family income in 1971 (%)	<$1000	24 (2.1)	5 (1.2)
	$1000–1999	47 (4.0)	13 (3.2)
	$2000–2999	52 (4.5)	16 (4.0)
	$3000–3999	44 (3.8)	19 (4.7)

	level	No	Yes
	$4000–4999	67 (5.8)	16 (4.0)
	$5000–5999	65 (5.6)	13 (3.2)
	$6000–6999	46 (4.0)	19 (4.7)
	$7000–7999	202 (17.4)	82 (20.3)
	$8000–9999	295 (25.3)	122 (30.3)
	$10000–14999	172 (14.8)	48 (11.9)
	$15000–24999	87 (7.5)	27 (6.7)
	$25000+	63 (5.4)	23 (5.7)
Marital status in 1971 (%)	Under 17	0 (0.0)	0 (0.0)
	Married	933 (77.7)	346 (80.8)
	Widowed	70 (5.8)	27 (6.3)
	Never married	76 (6.3)	20 (4.7)
	Divorced	77 (6.4)	22 (5.1)
	Separated	45 (3.7)	12 (2.8)
	Unknown	0 (0.0)	1 (0.2)
Highest grade of regular school ever in 1971 (mean (SD))		11.12 (3.00)	11.17 (3.32)
Amount of education in 1971 (%)	8th grade or less	218 (18.2)	93 (21.7)
	High School dropout	273 (22.7)	78 (18.2)
	High School	495 (41.2)	164 (38.3)
	College dropout	96 (8.0)	30 (7.0)
	College or more	119 (9.9)	63 (14.7)
Height in centimeters in 1971 (mean (SD))		168.48 (9.00)	169.48 (9.17)
Weight in kilograms in 1971 (mean (SD))		70.49 (15.57)	72.63 (16.08)
Weight change in kilograms (mean (SD))		1.98 (7.45)	4.53 (8.75)
Systolic blood pressure in 1982 (mean (SD))		127.70 (18.77)	131.69 (19.57)
Diastolic blood pressure in 1982 (mean (SD))		77.36 (10.54)	78.88 (10.84)
CHECK STATE CODE - SECOND PAGE (mean (SD))		31.95 (14.53)	30.58 (14.39)
Number of cigarettes smoked per day in 1971 (mean (SD))		21.18 (11.58)	18.79 (12.26)
YEARS OF SMOKING (mean (SD))		24.25 (11.83)	26.61 (13.03)
Diagnosed asthma in 1971 (%)	Never	1149 (95.7)	401 (93.7)
	Ever	52 (4.3)	27 (6.3)
Diagnosed chronic bronchitis/emphysema in 1971 (%)	Never	1101 (91.7)	389 (90.9)
	Ever	100 (8.3)	39 (9.1)
Diagnosed tuberculosis in 1971 (%)	Never	1188 (98.9)	418 (97.7)
	Ever	13 (1.1)	10 (2.3)
Diagnosed heart failure in 1971 (%)	Never	1195 (99.5)	426 (99.5)
	Ever	6 (0.5)	2 (0.5)
Diagnosed high blood pressure in 1971 (%)	Never	524 (43.6)	184 (43.0)
	Ever	88 (7.3)	42 (9.8)
	Missing	589 (49.0)	202 (47.2)
Diagnosed peptic ulcer in 1971 (%)	Never	1083 (90.2)	377 (88.1)
	Ever	118 (9.8)	51 (11.9)
Diagnosed colitis in 1971 (%)	Never	1156 (96.3)	418 (97.7)
	Ever	45 (3.7)	10 (2.3)

	level	No	Yes
Diagnosed hepatitis in 1971 (%)	Never	1179 (98.2)	422 (98.6)
	Ever	22 (1.8)	6 (1.4)
Diagnosed chronic cough in 1971 (%)	Never	1140 (94.9)	401 (93.7)
	Ever	61 (5.1)	27 (6.3)
Diagnosed hay fever in 1971 (%)	Never	1096 (91.3)	387 (90.4)
	Ever	105 (8.7)	41 (9.6)
Diagnosed diabetes in 1971 (%)	Never	600 (50.0)	224 (52.3)
	Ever	12 (1.0)	2 (0.5)
	Missing	589 (49.0)	202 (47.2)
Diagnosed polio in 1971 (%)	Never	1186 (98.8)	420 (98.1)
	Ever	15 (1.2)	8 (1.9)
Diagnosed malignant tumor/growth in 1971 (%)	Never	1172 (97.6)	419 (97.9)
	Ever	29 (2.4)	9 (2.1)
Diagnosed nervous breakdown in 1971 (%)	Never	1163 (96.8)	419 (97.9)
	Ever	38 (3.2)	9 (2.1)
Have you had 1 drink past year? (%)	Never	146 (12.2)	61 (14.3)
	Ever	1051 (87.5)	366 (85.5)
	Missing	4 (0.3)	1 (0.2)
How often do you drink? (%)	Almost every day	247 (23.5)	89 (24.3)
	2-3 times/week	179 (17.0)	52 (14.2)
	1-4 times/month	365 (34.7)	141 (38.5)
	<12 times/year	260 (24.7)	84 (23.0)
Which do you most frequently drink? (%)	Beer	429 (35.7)	140 (32.7)
	Wine	96 (8.0)	36 (8.4)
	Liquor	372 (31.0)	140 (32.7)
	Other/Unknown	304 (25.3)	112 (26.2)
When you drink, how much do you drink? (mean (SD))		3.35 (3.13)	3.09 (2.53)
Do you eat dirt or clay, starch or other non-standard food? (%)	Never	606 (50.5)	225 (52.6)
	Ever	6 (0.5)	1 (0.2)
	Missing	589 (49.0)	202 (47.2)
Use headache medication in 1971 (%)	Never	434 (36.1)	169 (39.5)
	Ever	767 (63.9)	259 (60.5)
Use other pains medication in 1971 (%)	Never	905 (75.4)	323 (75.5)
	Ever	296 (24.6)	105 (24.5)
Use weak heart medication in 1971 (%)	Never	1173 (97.7)	420 (98.1)
	Ever	28 (2.3)	8 (1.9)
Use allergies medication in 1971 (%)	Never	1129 (94.0)	399 (93.2)
	Ever	72 (6.0)	29 (6.8)
Use nerves medication in 1971 (%)	Never	1021 (85.0)	373 (87.1)
	Ever	180 (15.0)	55 (12.9)
Use lack of pep medication in 1971 (%)	Never	1136 (94.6)	410 (95.8)
	Ever	65 (5.4)	18 (4.2)
Use high blood pressure medication in 1971 (%)	Never	578 (48.1)	204 (47.7)
	Ever	34 (2.8)	22 (5.1)
	Missing	589 (49.0)	202 (47.2)
Use bowel trouble medication in 1971 (%)	Never	536 (44.6)	198 (46.3)
	Ever	76 (6.3)	28 (6.5)
	Missing	589 (49.0)	202 (47.2)
Use weight loss medication in 1971 (%)	Never	1167 (97.2)	420 (98.1)
	Ever	34 (2.8)	8 (1.9)
Use infection medication in 1971 (%)	Never	1019 (84.8)	369 (86.2)

	level	No	Yes
	Ever	182 (15.2)	59 (13.8)
In your usual day, how active are you? (%)	Very active	547 (45.5)	182 (42.5)
	Moderately active	540 (45.0)	198 (46.3)
	Inactive	114 (9.5)	48 (11.2)
In recreation, how much exercise? (%)	Much exercise	247 (20.6)	70 (16.4)
	Moderate exercise	496 (41.3)	181 (42.3)
	Little or no exercise	458 (38.1)	177 (41.4)
Birth control pills past 6 months? (mean (SD))		1.04 (0.95)	1.20 (0.94)
Total number of pregnancies? (mean (SD))		3.67 (2.20)	3.78 (2.22)
Serum cholesterol (mg/100ml) in 1971 (mean (SD))		218.90 (45.13)	222.96 (46.22)
AVG TOBACCO PRICE IN STATE OF RESIDENCE 1971 (US$2008) (mean (SD))		2.14 (0.23)	2.14 (0.23)
TOBACCO TAX IN STATE OF RESIDENCE 1971 (US$2008) (mean (SD))		1.06 (0.21)	1.06 (0.22)

Estimating the Unadjusted Average Treatment Effect (ATE) of Smoking Cessation on Health Outcomes

In this section, we estimate the **unadjusted Average Treatment Effect (ATE)** of smoking cessation on three key health outcomes:

- **Weight change (wt82_71)** between 1971 and 1982
- **Systolic blood pressure (sbp)** in 1982
- **Diastolic blood pressure (dbp)** in 1982

The **ATE** is calculated as the difference in the mean outcome between individuals who **quit smoking** (treatment group) and those who **continued smoking** (control group). The standard error (SE) is computed, and a **95% confidence interval (CI)** is constructed for each outcome.

Table 2: ATE Estimates for Smoking Cessation on Health Outcomes

Outcome	ATE	Standard.Error	Confidence.Interval
Weight Change (wt82_71)	2.541	0.474	(1.611, 3.47)
Systolic Blood Pressure (sbp)	3.991	1.090	(1.854, 6.128)
Diastolic Blood Pressure (dbp)	1.525	0.606	(0.338, 2.713)

Estimating the Average Treatment Effect (ATE) for Legislator Response in the Black Politicians Dataset

This section explores whether the **race of a legislator (Black or Non-Black)** influences their likelihood of responding to constituent emails. The dataset includes various district-level and legislator characteristics, with the primary treatment variable being **whether the legislator is Black (`leg_black`)** and the outcome variable being **whether they responded to the email (`responded`)**.

The first step in the analysis is to prepare the dataset by cleaning and transforming variables. Categorical variables, such as **legislator race (`leg_black`)** and **email response (`responded`)**, are converted into factor variables with meaningful labels (`Yes` or `No`). Other key categorical variables, including **whether the legislator is a senator (`leg_senator`)**, **party affiliation (`leg_democrat`)**, and **whether the legislator represents the southern U.S. (`south`)**, are also converted to factors. Additionally, variable labels are assigned to provide clear descriptions for each feature in the dataset.

Descriptive Statistics

To understand the distribution of key district and legislator characteristics, a **summary table** is created using the `tableone` package. This table presents means and proportions for different variables, stratified by legislator race. Importantly, **standardized mean differences (SMDs)** are included to measure the balance of covariates between Black and Non-Black legislators. Large SMD values indicate potential differences in baseline characteristics, which could be confounding factors in the analysis.

Table 3: Summary Statistics Stratified by Legislator Race

	level	No	Yes	p	test	SMD
n		5229	364			
Email is from out-of-district (mean (SD))		0.50 (0.50)	0.49 (0.50)	0.840		0.011
District Population (mean (SD))		8.42 (10.86)	10.43 (9.52)	0.001		0.197
District Median Household Income (mean (SD))		4.43 (1.40)	3.33 (1.14)	<0.001		0.867
District Median Household Income (Black) (mean (SD))		1.57 (1.02)	1.37 (0.43)	<0.001		0.259
District Median Household Income (White) (mean (SD))		2.33 (0.77)	2.22 (0.65)	0.009		0.152
Percentage of District that is Black (mean (SD))		0.06 (0.10)	0.52 (0.20)	<0.001		2.869
State's Squire Index (mean (SD))		0.19 (0.12)	0.21 (0.12)	0.001		0.175
Legislator is Neither Black nor White (mean (SD))		0.05 (0.22)	0.00 (0.00)	<0.001		0.326
Percentage of District that is Urban (mean (SD))		0.67 (0.32)	0.85 (0.25)	<0.001		0.630
Legislator is a Senator (%)	No	3839 (73.4)	278 (76.4)	0.240		0.068
	Yes	1390 (26.6)	86 (23.6)			
Legislator is in the Democratic Party (%)	No	2611 (49.9)	8 (2.2)	<0.001		1.296
	Yes	2618 (50.1)	356 (97.8)			
Legislator is in the Southern U.S. (%)	No	3920 (75.0)	168 (46.2)	<0.001		0.617
	Yes	1309 (25.0)	196 (53.8)			

Estimating the Unadjusted Average Treatment Effect (ATE)

The unadjusted ATE is computed as the difference in the mean email response rates between Black and Non-Black legislators. This calculation follows the standard ATE formula:

$ATE = E[Y|T = 1] - E[Y|T = 0]$

where (T=1) represents Black legislators and (T=0) represents Non-Black legislators. The mean email response rates are computed separately for each group, and their difference provides an initial estimate of the treatment effect.

Computing Standard Errors and Confidence Intervals

To assess the uncertainty of the ATE estimate, the standard error (SE) is calculated using the formula for a difference in proportions:

$SE = \sqrt{\frac{p_T(1-p_T)}{n_T} + \frac{p_C(1-p_C)}{n_C}}$

where ($ P_T$) and (p_C) are the response rates for Black and Non-Black legislators, respectively, and (n_T) and (n_C) are their corresponding sample sizes.

Using this standard error, a **95% confidence interval (CI)** is constructed to provide a range of plausible values for the ATE:

$CI = ATE \pm 1.96 \times SE$

If the confidence interval includes **zero**, it suggests that the difference in response rates between Black and Non-Black legislators **may not be statistically significant**.

Table 4: Unadjusted ATE Estimates for Legislator Response

Outcome	ATE	Standard_Error	X95..Confidence.Interval
Legislator Response	-0.032	0.026	[-0.084, 0.02]

Conclusion for the Analysis of Smoking Cessation and Legislator Response Patterns

Effect of Smoking Cessation on Health Outcomes

The analysis of the **National Health and Nutrition Examination Follow-up Study (NHEFS)** dataset revealed that quitting smoking had a significant impact on health outcomes. Using an **unadjusted ATE approach**, the following key results were observed:

- **Weight Change (1971–1982):** Those who quit smoking gained, on average, **2.54 kg** more weight compared to those who continued smoking. The **95% confidence interval (CI)** of **(1.61, 3.47)** indicates a statistically significant effect, confirming that smoking cessation is associated with weight gain.
- **Systolic Blood Pressure (SBP, 1982):** The estimated **ATE for SBP** was **3.99 mmHg**, with a **95% CI of (1.85, 6.13)**. This suggests that quitting smoking led to a modest increase in **systolic blood pressure**, which could have implications for cardiovascular risk management.
- **Diastolic Blood Pressure (DBP, 1982):** The **ATE for DBP** was **1.53 mmHg**, with a **95% CI of (0.34, 2.71)**. While smaller in magnitude compared to systolic pressure changes, this increase remains statistically significant.

These findings support existing literature suggesting that **smoking cessation is linked to weight gain**, potentially due to changes in metabolism and lifestyle habits. However, the **increase in blood pressure** warrants further investigation, as it may be influenced by weight gain and other behavioral changes post-cessation.

Effect of Legislator Race on Email Response Rates

The analysis of the **Black Politicians dataset** aimed to determine whether a legislator's race influences their likelihood of responding to constituent emails. The **unadjusted ATE estimation** yielded the following result:

- **Legislator Response Rate:** The **ATE was estimated at -0.032**, indicating that Black legislators were **3.2 percentage points less likely** to respond to constituent emails compared to Non-Black legislators. However, the **95% CI of (-0.084, 0.020)** includes **zero**, suggesting that this effect is **not statistically significant** at the 95% confidence level.

This finding suggests that, while there is a small observed difference in response rates, it is not large enough to rule out random variation. Further analyses using **adjusted methods, propensity score techniques, or subgroup analyses** could help clarify whether other factors—such as district characteristics, party affiliation, or geographic location—may contribute to differences in responsiveness.

Key Takeaways & Next Steps

1. **For Smoking Cessation and Health Outcomes:**
   - The findings confirm that quitting smoking leads to **weight gain** and **modest increases in blood pressure**.
   - Future analyses should explore **adjusted causal estimates**, accounting for confounders like **physical activity, diet, and pre-existing health conditions**.
   - Investigating **long-term health impacts** beyond 1982 could provide deeper insights into the **trade-offs between smoking cessation and cardiovascular health**.
2. **For Legislator Response Analysis:**
   - The **racial disparity in legislator response rates was small and statistically insignificant** in the unadjusted analysis.
   - Additional research should **adjust for potential confounders** like **district demographics, party affiliation, and political competition**.
   - Exploring **interaction effects** (e.g., how district-level racial composition influences legislator response behavior) could provide further insights.