Causal Inference 1: Police Stops

Eric Chan

2023-02-01

Table of contents

Import File 1

Does race correlate with REASON for traffic stop? 1
- Linear Probability Models (LPMs) . . . 2
- Logistic Regression Models . . . 2

Does race correlate with OUTCOMES of traffic stops? 3
- Linear Probability Models . . . 3
- Logistic Regression Models . . . 4

How do demographics of whose who receive a citation differ from those who do not recieve a citation? 4
- Linear Probability Models . . . 5
- Logistic Regression Models . . . 5

Import File

```r
# Import Data
nashville <- read.csv("nashville.csv")
```

Does race correlate with REASON for traffic stop?

Let's take a look at how race correlates with REASONS for traffic stops, such as moving violations, equipment issues, safety issues, seatbelts, and investigations.

Table 1: Table 1. Police Stops: Is race predictive of reasons for stop? (LPMs)

	moving	equipment	safety	seatbelt	investigate
(Intercept)	0.534***	0.304***	0.055***	0.030***	0.017***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)
black	−0.071***	0.043***	0.015***	0.001	0.004*
	(0.007)	(0.007)	(0.003)	(0.003)	(0.002)
Num.Obs.	19 942	19 942	19 942	19 942	19 942
R2	0.005	0.002	0.001	0.000	0.000
R2 Adj.	0.005	0.002	0.001	0.000	0.000
AIC	28 855.4	26 153.6	−397.0	−13 403.3	−23 027.0
BIC	28 879.1	26 177.3	−373.3	−13 379.6	−23 003.3
Log.Lik.	−14 424.692	−13 073.790	201.503	6704.658	11 516.498
F	93.954	39.457	19.552	0.137	4.696
RMSE	0.50	0.47	0.24	0.17	0.14

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Linear Probability Models (LPMs)

```
# Run models to test the differences of demographic variables.
library(modelsummary)
models_reasons <- list()

models_reasons[['moving']] <- lm(moving~black, data=nashville)
models_reasons[['equipment']] <- lm(equipment~black, data=nashville)
models_reasons[['safety']] <- lm(safety~black, data=nashville)
models_reasons[['seatbelt']] <- lm(seatbelt~black, data=nashville)
models_reasons[['investigate']] <- lm(investigate~black, data=nashville)

modelsummary(models_reasons, stars=TRUE, title="Table 1. Police Stops: Is race predictive
```

Logistic Regression Models

```
# Run models to test the differences of demographic variables.
library(modelsummary)
models_reasons <- list()

models_reasons[['moving']] <- glm(moving~black, data=nashville, family="binomial")
```

Table 2: Table 2. Police Stops: Is race predictive of reasons for stop? (Logistic)

	moving	equipment	safety	seatbelt	investigate
(Intercept)	0.137***	−0.828***	−2.838***	−3.460***	−4.047***
	(0.018)	(0.020)	(0.039)	(0.052)	(0.069)
black	−0.283***	0.195***	0.263***	0.031	0.228*
	(0.029)	(0.031)	(0.060)	(0.084)	(0.105)
Num.Obs.	19 942	19 942	19 942	19 942	19 942
AIC	27 551.6	24 973.7	9165.9	5493.8	3722.5
BIC	27 567.4	24 989.5	9181.7	5509.6	3738.3
Log.Lik.	−13 773.782	−12 484.874	−4580.942	−2744.911	−1859.263
F	93.340	39.336	19.442	0.137	4.677
RMSE	0.50	0.47	0.24	0.17	0.14

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
models_reasons[['equipment']] <- glm(equipment~black, data=nashville, family="binomial")
models_reasons[['safety']] <- glm(safety~black, data=nashville, family="binomial")
models_reasons[['seatbelt']] <- glm(seatbelt~black, data=nashville, family="binomial")
models_reasons[['investigate']] <- glm(investigate~black, data=nashville, family="binomial

modelsummary(models_reasons, stars=TRUE, title="Table 2. Police Stops: Is race predictive
```

Does race correlate with OUTCOMES of traffic stops?

Let's take a look at race and outcomes like frsiking, searching, warning, citation, and being arrested.

Linear Probability Models

```
models_outcomes <- list()
models_outcomes[['frisk']] <- lm(frisk~black, data=nashville)
models_outcomes[['search']] <- lm(search~black, data=nashville)
models_outcomes[['warning']] <- lm(warning~black, data=nashville)
models_outcomes[['citation']] <- lm(citation~black, data=nashville)
models_outcomes[['arrest']] <- lm(arrest~black, data=nashville)

modelsummary(models_outcomes, stars=TRUE, title="Table 3. Nashville Police Stops: Is race
```

Table 3: Table 3. Nashville Police Stops: Is race predictive of outcomes of stop? (LPM)

	frisk	search	warning	citation	arrest
(Intercept)	0.013***	0.030***	0.776***	0.239***	0.013***
	(0.001)	(0.002)	(0.004)	(0.004)	(0.001)
black	0.013***	0.026***	0.044***	−0.034***	0.008***
	(0.002)	(0.003)	(0.006)	(0.006)	(0.002)
Num.Obs.	19 989	19 989	19 987	19 987	19 989
R2	0.002	0.004	0.003	0.002	0.001
R2 Adj.	0.002	0.004	0.003	0.002	0.001
AIC	−23 486.1	−8810.4	20 598.0	21 843.4	−25 769.5
BIC	−23 462.4	−8786.7	20 621.7	21 867.1	−25 745.8
Log.Lik.	11 746.047	4408.181	−10 296.007	−10 918.696	12 887.765
F	47.120	81.330	54.474	31.407	17.873
RMSE	0.13	0.19	0.41	0.42	0.13

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Logistic Regression Models

```
models_outcomes <- list()
models_outcomes[['frisk']] <- glm(frisk~black, data=nashville, family="binomial")
models_outcomes[['search']] <- glm(search~black, data=nashville, family="binomial")
models_outcomes[['warning']] <- glm(warning~black, data=nashville, family="binomial")
models_outcomes[['citation']] <- glm(citation~black, data=nashville, family="binomial")
models_outcomes[['arrest']] <- glm(arrest~black, data=nashville, family="binomial")

modelsummary(models_outcomes, stars=TRUE, title="Table 4. Nashville Police Stops: Is race
```

How do demographics of whose who receive a citation differ from those who do not recieve a citation?

General demographics of whose who gets stopped, then receive a citation vs. not.

Table 4: Table 4. Nashville Police Stops: Is race predictive of outcomes of stop? (Logistic)

	frisk	search	warning	citation	arrest
(Intercept)	−4.302***	−3.486***	1.242***	−1.159***	−4.296***
	(0.078)	(0.053)	(0.022)	(0.021)	(0.078)
black	0.710***	0.647***	0.271***	−0.198***	0.467***
	(0.106)	(0.073)	(0.037)	(0.035)	(0.111)
Num.Obs.	19 989	19 989	19 987	19 987	19 989
AIC	3636.0	6559.7	20 365.2	21 329.3	3333.3
BIC	3651.8	6575.5	20 381.0	21 345.1	3349.1
Log.Lik.	−1816.005	−3277.846	−10 180.589	−10 662.638	−1664.671
F	45.192	78.561	54.147	31.308	17.553
RMSE	0.13	0.19	0.41	0.42	0.13

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Linear Probability Models

```
models_demo <- list()
models_demo[['Black']] <- lm(black~citation, data=nashville)
models_demo[['White']] <- lm(white~citation, data=nashville)
models_demo[['Hispanic']] <- lm(hispanic~citation, data=nashville)
models_demo[['Male']] <- lm(male~citation, data=nashville)
models_demo[['Age']] <- lm(subject_age~citation, data=nashville)

modelsummary(models_demo, stars=TRUE, title="Table 5. Demographics for receiving a citatio
```

Logistic Regression Models

```
models_demo <- list()
models_demo[['Black']] <- glm(black~citation, data=nashville, family="binomial")
models_demo[['White']] <- glm(white~citation, data=nashville, family="binomial")
models_demo[['Hispanic']] <- glm(hispanic~citation, data=nashville, family="binomial")
models_demo[['Male']] <- glm(male~citation, data=nashville, family="binomial")
models_demo[['Age']] <- lm(subject_age~citation, data=nashville)

modelsummary(models_demo, stars=TRUE, title="Table 6. Demographics for receiving a citatio
```

Table 5: Table 5. Demographics for receiving a citation (LPM, where applicable)

	Black	White	Hispanic	Male	Age
(Intercept)	0.389***	0.536***	0.048***	0.588***	37.553***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.114)
citation	−0.046***	0.004	0.037***	0.023**	−2.353***
	(0.008)	(0.008)	(0.004)	(0.008)	(0.239)
Num.Obs.	19 987	19 987	19 997	19 915	19 996
R2	0.002	0.000	0.005	0.000	0.005
R2 Adj.	0.002	0.000	0.005	0.000	0.005
AIC	27 773.9	28 907.9	−2025.8	28 208.3	162 732.6
BIC	27 797.6	28 931.6	−2002.1	28 232.0	162 756.3
Log.Lik.	−13 883.960	−14 450.947	1015.887	−14 101.149	−81 363.296
F	31.407	0.251	92.477	7.561	96.574
RMSE	0.48	0.50	0.23	0.49	14.15

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Table 6. Demographics for receiving a citation (Logistic, where applicable)

	Black	White	Hispanic	Male	Age
(Intercept)	−0.451***	0.145***	−2.990***	0.354***	37.553***
	(0.016)	(0.016)	(0.038)	(0.016)	(0.114)
citation	−0.198***	0.017	0.618***	0.095**	−2.353***
	(0.035)	(0.034)	(0.065)	(0.035)	(0.239)
Num.Obs.	19 987	19 987	19 997	19 915	19 996
R2					0.005
R2 Adj.					0.005
AIC	26 491.0	27 601.2	8586.3	26 914.2	162 732.6
BIC	26 506.8	27 617.0	8602.1	26 930.0	162 756.3
Log.Lik.	−13 243.494	−13 798.590	−4291.133	−13 455.081	−81 363.296
F	31.308	0.251	89.607	7.556	96.574
RMSE	0.48	0.50	0.23	0.49	14.15

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$