Ravi clas proj - Texas stops & race classification

tx1
csr315@nyu.edu mvb DUE: MONDAY MAY 11

dta: https://openpolicing.stanford.edu/data/
doc: https://github.com/stanford-policylab/opp/blob/master/data_readme.md code: https://github.com/stanford-policylab/opp/tree/master/lib

progs: 1 Looking at Texas state data.R 2 Downloading census data.R 3 Merging census data with Texas data 4 Clean Texas state data 5 Create descriptive tables 6 Run regressions

outline of report

clearly state and motivate your project/research question/policy evaluation idea, summarize related academic literature, describe your methods, detail your results (with the appropriate plots), and discuss the implications of your findings. Unless you receive permission for an exception from the instructor, you must also include a section discussing potential ethical implications or concerns related to the project .

outline of models to use

Models Predict misclassification: Outcomes Misclassification type H-W – white/other No Misclassification (H-H)

Covariates Year County Sex of subject Officer is hispanic Violation recode into categories (e.g. speeding) Speeding Stop sign Red light DUI Tail light or head lamp Med income age in geog level Med age in geog level %hisp in geog level *** %white in geog level %black in geog level Dummy majority nonwhite in geog level Dummy urban/nonwhite in geog level Maybe crime RS: maybe characteristics of stop - ..time of day /

Predict Hit rate: Outcomes Found contraband weapon/searches Found contraband drugs/searches Found contraband any/searches Covariates Same as above

Other ideas Veil of darkness RS: contraband found hit rates - does this vary by misc cat Demographics of person stopped varies by if state patrol data (city omre likely to be close to home) Discrepancies of hit rates - view all at once - see plots used in paper Points along 45 deg line along location. Sized by Frisky - plot 3A and plot B precinct or prejudice. X axis: proportion of residents that are hispanic. Y axis: misclassification rates. Prop classified as white. H: more hispanics, more misclassification rate

Misid outcome models

copy pasted from prog 4- Run Models

```
load("C:/Users/Claud/Box/ddpe/stx/dta/Texas1MSampleClean_Merged.Rdata")

# preprocessing ------------------------------------------------------------

tx_merged[mc!= "H-O", misid := 1*(mc == "H-W")]

tx<-tx_merged %>%
  filter(year(date)<2016 & year(date)> 2008 & mc %in% c("H-W", "H-H")) %>%
  mutate(year=year(date),misid=as.factor(misid))



# predict misid ------------------------------------------------------------
```

```r
model1 <-
  glm(
    misid ~ prop_black + prop_hisp +
      prop_white + prop_urban + median_age_dec_2010 +
      income_acs_2015 + subject_sex + year,
    data = tx,
    family = "binomial"
  )

model2 <-
  glm(
    misid ~ prop_black + prop_hisp +
      prop_white + prop_urban + median_age_dec_2010 +
      income_acs_2015 + subject_sex + year + officer_last_name_hisp,
    data = tx,
    family = "binomial"
  )


model3 <-
  glm(
    misid ~ prop_black + prop_hisp +
      prop_white + prop_urban + median_age_dec_2010 +
      income_acs_2015 + subject_sex + year + officer_last_name_hisp +
      viol_belt + viol_drug + viol_alcohol + viol_dui + viol_lamp +
      viol_license + viol_plate + viol_registration + viol_speed +
      viol_traffic + viol_mod,
    data = tx,
    family = "binomial"
  )

# vif(model3)
anova(model1, model2, test = "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: misid ~ prop_black + prop_hisp + prop_white + prop_urban + median_age_dec_2010 +
##     income_acs_2015 + subject_sex + year
## Model 2: misid ~ prop_black + prop_hisp + prop_white + prop_urban + median_age_dec_2010 +
##     income_acs_2015 + subject_sex + year + officer_last_name_hisp
##   Resid. Df Resid. Dev Df Deviance    Rao  Pr(>Chi)
## 1    238568     307265
## 2    238567     307198  1   66.811 66.889 2.872e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(model1, model2, model3, test = "Rao")
```

```
## Warning in if (df < 0) score[i + 1] <- -score[i + 1]: the condition has length >
## 1 and only the first element will be used

## Warning in if (df < 0) score[i + 1] <- -score[i + 1]: the condition has length >
## 1 and only the first element will be used

## Analysis of Deviance Table
```

```
##
## Model 1: misid ~ prop_black + prop_hisp + prop_white + prop_urban + median_age_dec_2010 +
##     income_acs_2015 + subject_sex + year
## Model 2: misid ~ prop_black + prop_hisp + prop_white + prop_urban + median_age_dec_2010 +
##     income_acs_2015 + subject_sex + year + officer_last_name_hisp
## Model 3: misid ~ prop_black + prop_hisp + prop_white + prop_urban + median_age_dec_2010 +
##     income_acs_2015 + subject_sex + year + officer_last_name_hisp +
##     viol_belt + viol_drug + viol_alcohol + viol_dui + viol_lamp +
##     viol_license + viol_plate + viol_registration + viol_speed +
##     viol_traffic + viol_mod
##   Resid. Df Resid. Dev Df Deviance    Rao  Pr(>Chi)
## 1    238568     307265
## 2    238567     307198  1    66.81  66.89 2.872e-16 ***
## 3    238556     306254 11   944.65 948.18 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(model1)
```

```
## [1] 307283.1
```

```
AIC(model2)
```

```
## [1] 307218.3
```

```
AIC(model3)
```

```
## [1] 306295.6
```

```
model3table <- tidy(model3)
write.csv(model3table, file = "logitmodelforppt.csv")
exp(coef(model3))
```

```
##            (Intercept)             prop_black              prop_hisp
##          6.210522e+235           2.327504e+00           2.536515e+00
##             prop_white             prop_urban    median_age_dec_2010
##           3.142136e+00           9.507940e-01           1.007292e+00
##        income_acs_2015      subject_sexfemale                   year
##           1.000002e+00           1.272207e+00           7.630995e-01
## officer_last_name_hisp              viol_belt             viol_drug
##           9.178115e-01           9.227280e-01           9.706111e-01
##           viol_alcohol               viol_dui             viol_lamp
##           8.774268e-01           9.410187e-01           9.191074e-01
##           viol_license             viol_plate      viol_registration
##           7.148632e-01           1.135504e+00           1.121692e+00
##             viol_speed           viol_traffic               viol_mod
##           9.924194e-01           9.887063e-01           1.032234e+00
```

```
#stargazer(tx[1:3], summary=FALSE, header=FALSE,
          #digits=2,
         # title="Descriptives table")

stargazer(data=model1, model2, model3,
          header=FALSE,
          type='latex',
          title            = "Logistic Reg on Misid",
          dep.var.caption  = "Hispanic Driver Recorded as White",
           column.labels = c("Model1", "Model2", "Model3"))
```

Including Plots

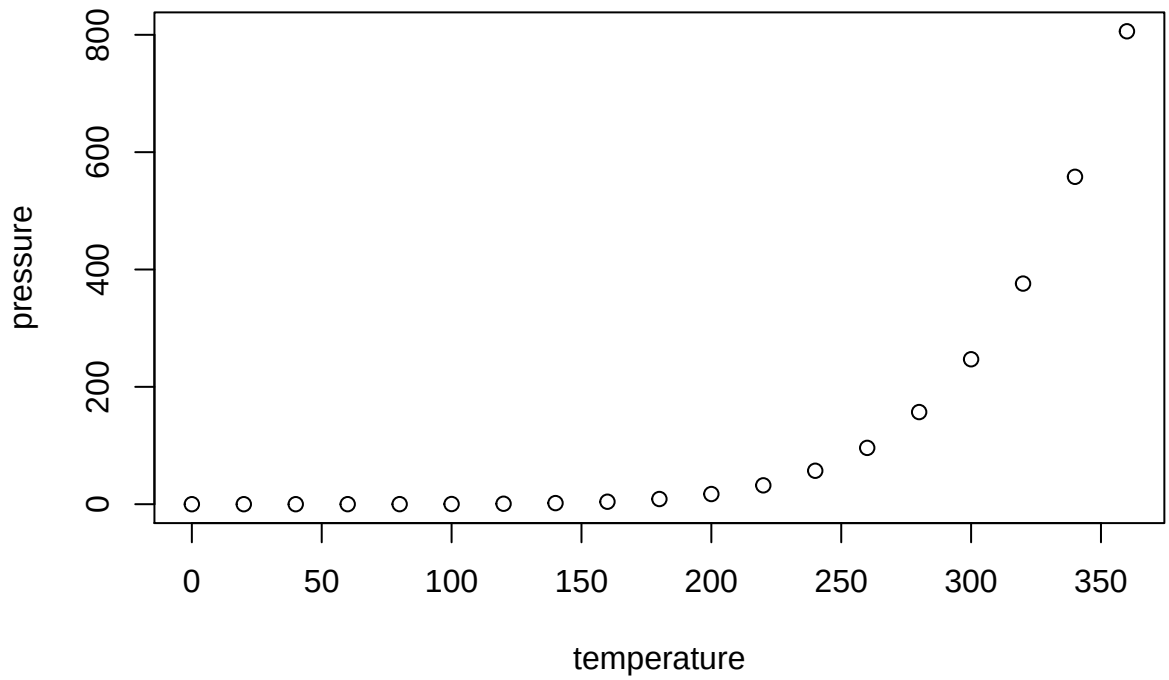

Table 1: Logistic Reg on Misid

	Hispanic Driver Recorded as White		
	misid		
	Model1	Model2	Model3
	(1)	(2)	(3)
prop_black	0.793***	0.766***	0.845***
	(0.189)	(0.189)	(0.189)
prop_hisp	0.883***	0.924***	0.931***
	(0.163)	(0.164)	(0.164)
prop_white	1.089***	1.085***	1.145***
	(0.166)	(0.166)	(0.166)
prop_urban	−0.079***	−0.078***	−0.050***
	(0.014)	(0.014)	(0.014)
median_age_dec_2010	0.008***	0.008***	0.007***
	(0.001)	(0.001)	(0.001)
income_acs_2015	0.00000***	0.00000***	0.00000***
	(0.00000)	(0.00000)	(0.00000)
subject_sexfemale	0.241***	0.242***	0.241***
	(0.010)	(0.010)	(0.010)
year	−0.272***	−0.272***	−0.270***
	(0.002)	(0.002)	(0.002)
officer_last_name_hisp		−0.085***	−0.086***
		(0.010)	(0.010)
viol_belt			−0.080***
			(0.020)
viol_drug			−0.030
			(0.051)
viol_alcohol			−0.131**
			(0.051)
viol_dui			−0.061
			(0.041)
viol_lamp			−0.084***
			(0.015)
viol_license			−0.336***
			(0.012)
viol_plate			0.127***
			(0.013)
viol_registration			0.115***
			(0.013)
viol_speed			−0.008
			(0.012)