Final Take Home Exam 2022

## Part 1: Determinants of Voter Turnout

#### What is the null hypothesis for the β3 term in model 2?

The null hypothesis for β3 is that the Gini index of disposable income does not have an effect on voter turnout.

#### 2) Interpret the effects of electoral competition (i.e. a change in government) on turnout using the coefficients from models 1 and 2.

The coefficient change in government is positive and statistically significant to a 2.02 increase in voter turnout in model 1. In model 2 a change in government has a 2.99 increase but it is not statistically significant.

#### 3) Based on model 2, what is the expected level of turnout for a country with the following characteristics?

##### • A) A county with a new government, with a level of expenditure in education of 4% of the GDP, with a GINI of 40% and an effective number of parties of 5

The level of voter turnout would be 54.04.

##### • B) A county with an old government, with a level of expenditure in education of 4% of the GDP, with a GINI of 40% and an effective number of parties of 5

The level of voter turnout would be 51.05.

##### • C) A county with a new government, with a level of expenditure in education of 8% of the GDP, with a GINI of 40% and an effective number of parties of 5

The level of voter turnout would be 54.08.

##### • D) A county with an old government, with a level of expenditure in education of 8% of the GDP, with a GINI of 40% and an effective number of parties of 5

The level of voter turnout would be 51.09.

## Part 2: Predicting the Black Live Matters protests

## `geom\_smooth()` using formula 'y ~ x'

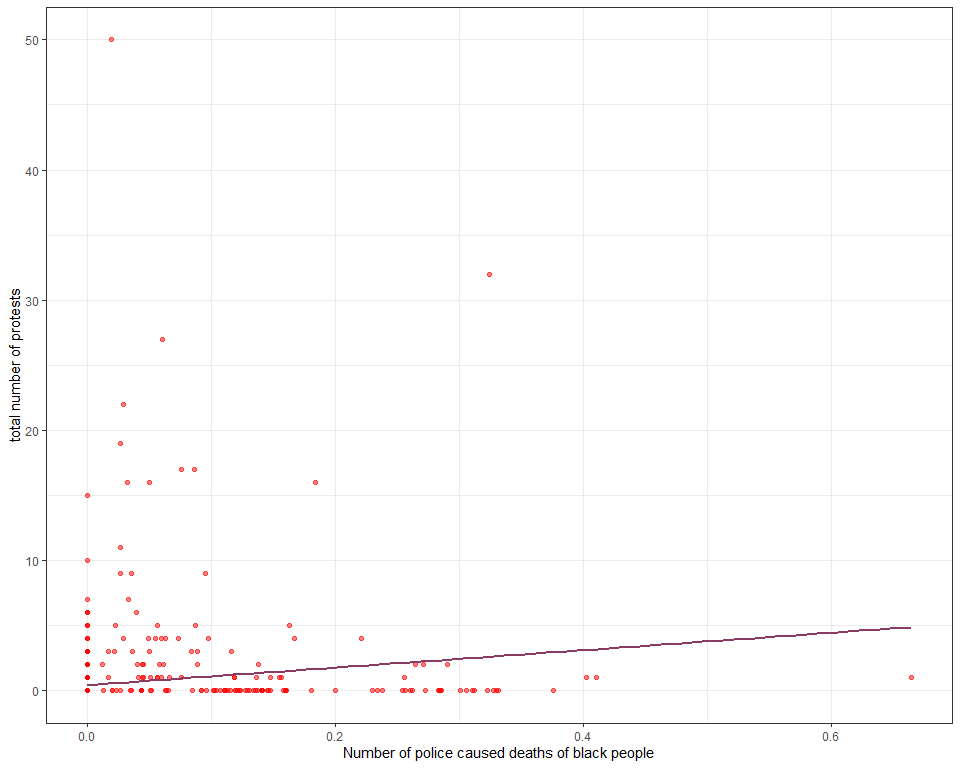


Figure 1.

We can see from the regression line that it is upward sloping meaning that the higher number of protests are related to higher numbers of police caused deaths of black people. From the regression output below we can see that the coeficient for the variable deaths\_black\_pc meaning the number of police caused deaths of black people is just above 0, the estimated intercept coefficient is 0.44

screenreg(list(model1, multi\_model\_1))

##   
## =========================================  
## Model 1 Model 2   
## -----------------------------------------  
## (Intercept) 0.44 \*\*\* -1.40 \*\*\*  
## (0.09) (0.26)   
## deaths\_black\_pc 6.70 \*\*\* 2.91 \*\*   
## (1.38) (0.93)   
## crime 1.05 \*\*\*  
## (0.21)   
## Per\_Black 0.01   
## (0.00)   
## TotalPop 0.00 \*\*\*  
## (0.00)   
## mayorrep -0.00   
## (0.12)   
## blackmayor -0.45 \*   
## (0.22)   
## dem\_share 0.02 \*\*\*  
## (0.00)   
## -----------------------------------------  
## R^2 0.02 0.61   
## Adj. R^2 0.02 0.61   
## Num. obs. 1066 1066   
## =========================================  
## \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

For model 2, I chose 6 additional variables alongside deaths\_black\_pc, these were if the city is listed in the top 100 highest violent crime cities (crime). My hypothesis is that a city with a higher crime rate will have a higher number of black people killed by the police, in turn having a higer level of protests. The percentage of black people in the total poulation of a city. My hypothesis is that cities with higher percentages of black people living in the city will have more protests (Per\_Black).The total population of the locality. my hypothesis is that larger population centers will have more protests (TotalPop). If a city has a republican mayor, my hypothesis is that cities with republican mayors will have higher numbers of protests (mayorrep), if a city has a black mayor. my hypothesis is that a city with a black mayor will have fewer protests (blackmayor) this is also my hypothesis for the democratic vote in the 2008 presidential election (dem\_share). The null hypothesis is that the vaibles will not have an effect on the number of protests in a city.

From the regression output above, we can see that the coefficient deaths\_black\_pc there is a 6.70 increase on protests with a p value below 0.001 making it statiscally significant. In model 2, with the other variables the deaths has a decrease of 2.91. The percentage of black people, the total population and if there is a republican mayor do not have an effect on the total number of protests, so I cannot reject the null hypothesis.

If a city has a black mayor has a deacrese effect on the total number of protests by - 0.45 and a p value below 0.05 making it statistically significant. The R squared for model 2 (0.66) is higher than model 1 (0.02) , indicating that model 2 has a better fit.

It is clear that the increase in police caused deaths of black people have increse the number of protests while other varibles have an incresing effect such as crime rate.

## Code:

install.packages("texreg") # install only once  
library(texreg) # load in the beginning of every R session  
library(tidyverse) # also load tidyverse!  
library(ggplot2)  
  
# Dataset  
blm\_dat <- read.csv("https://raw.githubusercontent.com/QMUL-SPIR/Public\_files/master/datasets/BLM\_exam.csv")  
  
# Part 2: Predicting the Black Live Matters protests  
  
# Descriptive statistics  
  
# table of regression output  
  
### model 1 ###  
# dependent variable = tot.protests   
# explanatory variable = deaths\_black\_pc  
  
  
  
# summary of varibles   
summary(blm\_dat$tot.protests)  
summary(blm\_dat$deaths\_black\_pc)  
  
  
# scatterplot   
p1 <- ggplot(blm\_dat, aes(y = tot.protests, x = deaths\_black\_pc)) + # basic plot instructions  
 geom\_point( # make it a scatterplot  
 color = "red",   
 alpha = 0.5, # make the points semi-transparent  
 shape = 19) + # make the points circles  
 labs(x = "Number of police caused deaths of black people", y = "total number of protests") + # axis labels  
 theme\_bw() # simplistic theme  
  
p1  
  
# regression  
  
model1 <- lm(tot.protests ~ deaths\_black\_pc, # formula  
 data = blm\_dat) # dataset  
summary(model1)  
  
  
# add regression line  
  
# add the geom\_smooth() funciton to add the regression line  
p1 + geom\_smooth(method = "lm", # line should be fit as a linear model  
 color = "hotpink4",   
 se = FALSE) # hide the confidence intervals  
  
# present finding in an easy to read table  
screenreg(model1)  
  
### model 2 ###  
  
summary(blm\_dat)  
  
# dependent variable = tot.protests   
# explanatory variable = deaths\_black\_pc, crime, Per\_Black, TotalPop, mayorrep, blackmayor , dem\_share  
  
# model with more explanatory variables  
multi\_model\_1 <- lm(tot.protests ~   
 deaths\_black\_pc + crime + Per\_Black + TotalPop + mayorrep + blackmayor + dem\_share,  
 data = blm\_dat)  
  
summary(multi\_model\_1)  
  
screenreg(list(model1, multi\_model\_1))