

Coding_Sample_2

2023-10-24

```
# load packages
library(dplyr)
library(tidyverse)
library(stargazer)
library(sandwich)
library(stringr)

# load datafile
data1 <- read.csv("/Users/martrinmunoz/Desktop/EconPredoc/Writing Samples/Cocaine Project/col_coc.csv")
# summarize the data
summary(data1)
```

```
##      year      coca.production..ha. eradication..ha.
## Min.   :2002   Min.    :    0.0      Min.    :    0
## 1st Qu.:2005   1st Qu.:  133.5      1st Qu.:    0
## Median :2009   Median :   994.0      Median :   47
## Mean   :2009   Mean    : 3520.8      Mean    : 4420
## 3rd Qu.:2013   3rd Qu.: 4816.2      3rd Qu.: 4627
## Max.   :2016   Max.    :45516.0      Max.    :71891
## europe.cocaine.wholesale.prices..US..per.kg.inflation.adjusted.to.20xx.euros.and.purity.adjusted.
## Length:360
## Class :character
## Mode  :character
##
##
##
## agriculture.livestock.price.index..2004.2006... unemployment
## Min.    : 90.85                                Min.    :0.0570
## 1st Qu.: 98.82                                1st Qu.:0.0920
## Median :103.37                                Median :0.1080
## Mean    :105.33                                Mean    :0.1091
## 3rd Qu.:114.39                                3rd Qu.:0.1220
## Max.    :119.47                                Max.    :0.2230
##      X2003      X2004      X2005      X2006
## Min.    :0.00000  Min.    :0.00000  Min.    :0.00000  Min.    :0.00000
## 1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.00000
## Median :0.00000  Median :0.00000  Median :0.00000  Median :0.00000
## Mean    :0.06667  Mean    :0.06667  Mean    :0.06667  Mean    :0.06667
## 3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:0.00000
## Max.    :1.00000  Max.    :1.00000  Max.    :1.00000  Max.    :1.00000
##      X2007      X2008      X2009      X2010
## Min.    :0.00000  Min.    :0.00000  Min.    :0.00000  Min.    :0.00000
## 1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.00000
## Median :0.00000  Median :0.00000  Median :0.00000  Median :0.00000
## Mean    :0.06667  Mean    :0.06667  Mean    :0.06667  Mean    :0.06667
```

##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	X2011	X2012	X2013	X2014
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.06667	Mean :0.06667	Mean :0.06667	Mean :0.06667
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	X2015	X2016	antioquia	arauca
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.06667	Mean :0.06667	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	bolivar	boyaca	caldas	caqueta
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.04167	Mean :0.04167	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	cauca	cesar	choco	cordoba
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.04167	Mean :0.04167	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	cundinamarca	guainia	guaviare	guajira
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.04167	Mean :0.04167	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	magdalena	meta	narino	nortedesantander
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.04167	Mean :0.04167	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	putumayo	santander	valledelcauca	vaupes
##	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :0.00000	Median :0.00000
##	Mean :0.04167	Mean :0.04167	Mean :0.04167	Mean :0.04167
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.00000
##	vichada			
##	Min. :0.00000			
##	1st Qu.:0.00000			

```

## Median :0.00000
## Mean   :0.04167
## 3rd Qu.:0.00000
## Max.   :1.00000

# let's rename some of the columns
data1 <- data1 %>%
  rename(
    coca_prod = coca.production..ha.,
    eradication = eradication..ha.,
    price = europe.cocaine.wholesale.prices..US..per.kg.inflation.adjusted.to.20xx.euros.and.purity.adjusted.to.2004.2006...
  )
# The year variables have an 'X' in front of them. Let's remove these.
for (i in 7:20){
  colnames(data1)[i] <- str_sub(colnames(data1)[i], 2, 5)
}
# identify columns that contain non-integer or non-numeric variables
data1 %>%
  select_if(function(x) !is.numeric(x) & !is.integer(x)) %>%
  colnames(.)

## [1] "price"

# let's print this column to see what is the problem
head(data1[, "price"])

## [1] "$51.07" "$51.07" "$51.07" "$51.07" "$51.07" "$51.07"

# The price variable has a '$' sign in front of each dollar amount that needs
# to be removed. Let's remove it now
data1 <- data1 %>%
  mutate(price = str_replace(price, "\\$", ""))
# Now we can convert price to a numeric variable
data1$price <- as.numeric(data1$price)
# Make sure this worked
class(data1$price)

## [1] "numeric"

# Let's run a linear regression of coca production on all the variables except for
# price, agriculture, and year
model1 <- lm(coca_prod ~ . - price - agriculture - year, data = data1)
summary(model1)

##
## Call:
## lm(formula = coca_prod ~ . - price - agriculture - year, data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##    -8101   -1102     -19     828   36408
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.083e+02  2.020e+03   0.103  0.917911
## eradication    -1.727e-02  3.064e-02  -0.564  0.573431

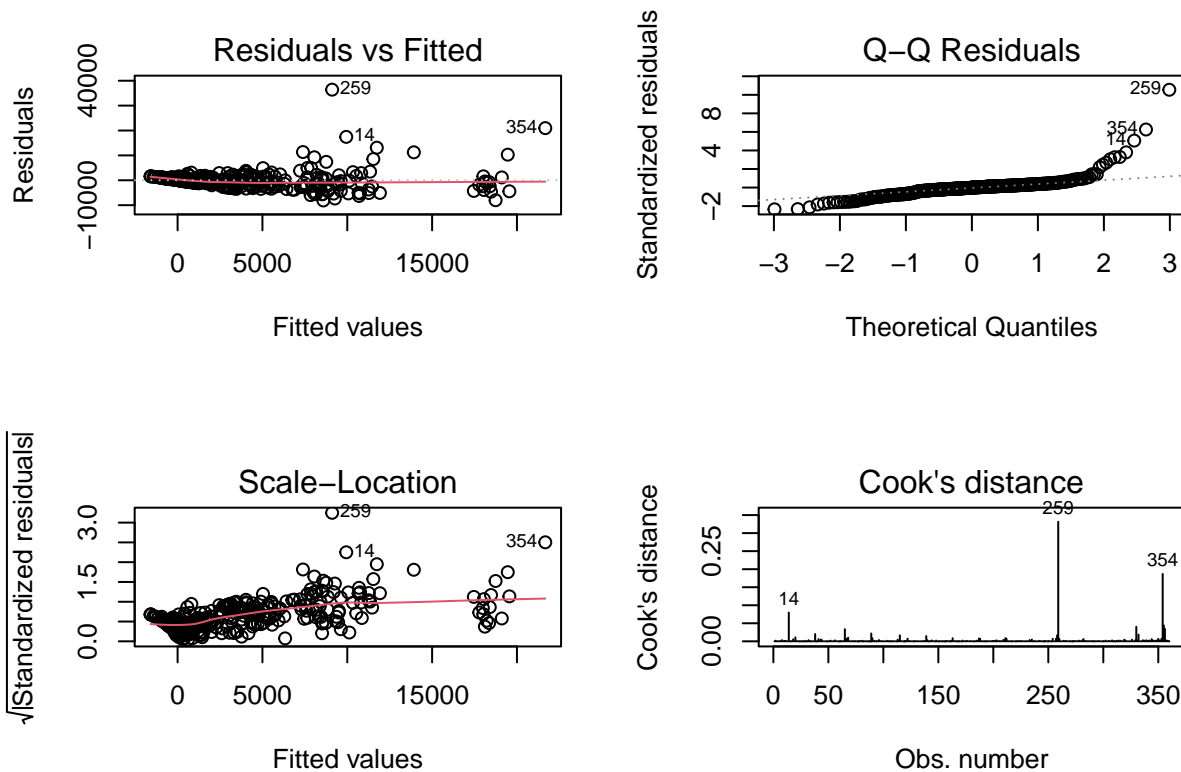
```

```

## unemployment      6.872e+03  1.147e+04   0.599 0.549492
## `2003`             -5.601e+02  1.066e+03  -0.525 0.599701
## `2004`             -9.299e+02  1.072e+03  -0.867 0.386567
## `2005`             -1.437e+02  1.113e+03  -0.129 0.897389
## `2006`             -8.352e+02  1.084e+03  -0.771 0.441390
## `2007`              4.142e+01  1.088e+03   0.038 0.969644
## `2008`             -7.033e+02  1.094e+03  -0.643 0.520792
## `2009`             -1.090e+03  1.083e+03  -1.007 0.314662
## `2010`             -1.541e+03  1.086e+03  -1.419 0.156906
## `2011`             -1.454e+03  1.114e+03  -1.305 0.192707
## `2012`             -3.254e+02  1.133e+03  -0.287 0.774149
## `2013`             -2.011e+03  1.164e+03  -1.727 0.085136 .
## `2014`             -1.097e+03  1.189e+03  -0.922 0.357152
## `2015`              6.768e+01  1.203e+03   0.056 0.955171
## `2016`              2.293e+03  1.186e+03   1.934 0.053939 .
## antioquia          4.481e+03  1.361e+03   3.294 0.001099 **
## arauca              3.285e+02  1.335e+03   0.246 0.805763
## bolivar            2.791e+03  1.366e+03   2.043 0.041887 *
## boyaca             -1.429e+02  1.349e+03  -0.106 0.915710
## caldas             -3.953e+02  1.338e+03  -0.295 0.767896
## caqueta            5.515e+03  1.404e+03   3.927 0.000105 ***
## cauca              4.609e+03  1.352e+03   3.408 0.000737 ***
## cesar              -3.429e+02  1.342e+03  -0.255 0.798544
## choco              1.228e+03  1.337e+03   0.919 0.358915
## cordoba            1.216e+03  1.336e+03   0.911 0.363189
## cundinamarca       -3.674e+02  1.335e+03  -0.275 0.783344
## guainia            1.580e+01  1.334e+03   0.012 0.990556
## guaviare           8.790e+03  1.392e+03   6.314 9.07e-10 ***
## guajira            -9.812e+01  1.370e+03  -0.072 0.942964
## magdalena          -2.840e+01  1.367e+03  -0.021 0.983436
## meta              7.510e+03  1.354e+03   5.546 6.12e-08 ***
## narino             1.865e+04  1.632e+03  11.426 < 2e-16 ***
## nortedesantander   8.373e+03  1.341e+03   6.244 1.36e-09 ***
## putumayo           1.079e+04  1.417e+03   7.615 3.00e-13 ***
## santander          3.335e+02  1.347e+03   0.248 0.804543
## valledelcauca      2.571e+01  1.347e+03   0.019 0.984783
## vaupes             1.739e+02  1.334e+03   0.130 0.896335
## vichada            3.152e+03  1.335e+03   2.362 0.018785 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3653 on 320 degrees of freedom
## Multiple R-squared:  0.6461, Adjusted R-squared:  0.603
## F-statistic: 14.98 on 39 and 320 DF,  p-value: < 2.2e-16

# get model diagnostics
par(mfrow=c(2,2))
plot(model1, which=1:4)

```



```
# Notice that the standard errors are a bit heteroskedastic since we don't get a
# perfectly straight line in the bottom left graph
# Let's get heteroskedastic robust errors for our model
```

```
se1 <- vcovHC(model1, type = "HC1")
robust_se1 <- sqrt(diag(se1))
```

```
# Now let's report the linear regression in a nice table. We want this table to
# capitalize the first letter of the names of the independent variables.
# I will do this manually for the first few independent variables, but will
# use a specific function to capitalize the first letter of all entity fixed
# effects in one go. I will first collect the entity fixed effects into a vector
# then I will apply the str_to_title function to capitalize the first letter of
# each element of the vector.
```

```
departments <- colnames(data1[,21:ncol(data1)])
departments <- str_to_title(departments)
```

```
# Now let's report the linear regression in a table. Note that the first three
# independent variables are manually renamed, the time fixed effects are renamed
# using the rep function, and the entity fixed effects are renamed using the
# character vector created in the previous code chunk. Also notice that the
# intercept is moved to the top of the table using the intercept.bottom argument.
```

```
stargazer(model1, type = 'latex', title = "OLS Regression: The effects of aerial
spraying on coca production", header=FALSE, se = list(NULL, robust_se1),
digits = 2, covariate.labels= c("Intercept", "Eradication", "Unemployment",
rep(2003:2016), departments),
intercept.bottom = FALSE, single.row = TRUE,
dep.var.labels = "Coca Production")
```

Table 1: OLS Regression: The effects of aerial spraying on coca production

	<i>Dependent variable:</i>
	Coca Production
Intercept	208.33 (2,019.72)
Eradication	-0.02 (0.03)
Unemployment	6,872.30 (11,469.95)
2003	-560.06 (1,066.06)
2004	-929.87 (1,072.46)
2005	-143.71 (1,113.49)
2006	-835.22 (1,083.57)
2007	41.42 (1,087.65)
2008	-703.28 (1,094.03)
2009	-1,090.23 (1,082.58)
2010	-1,541.28 (1,086.26)
2011	-1,453.68 (1,113.62)
2012	-325.42 (1,133.10)
2013	-2,010.67* (1,164.27)
2014	-1,096.72 (1,189.33)
2015	67.68 (1,202.97)
2016	2,293.49* (1,185.60)
Antioquia	4,481.16*** (1,360.51)
Arauca	328.52 (1,334.91)
Bolivar	2,791.35** (1,366.40)
Boyaca	-142.90 (1,349.10)
Caldas	-395.34 (1,338.42)
Caqueta	5,515.35*** (1,404.45)
Cauca	4,608.62*** (1,352.19)
Cesar	-342.87 (1,342.24)
Choco	1,227.93 (1,336.52)
Cordoba	1,216.36 (1,335.77)
Cundinamarca	-367.42 (1,335.11)
Guainia	15.80 (1,333.85)
Guaviare	8,790.12*** (1,392.14)
Guajira	-98.12 (1,370.42)
Magdalena	-28.40 (1,366.86)
Meta	7,509.88*** (1,354.09)
Narino	18,645.82*** (1,631.83)
Nortedesantander	8,372.61*** (1,340.97)
Putumayo	10,788.60*** (1,416.79)
Santander	333.52 (1,346.59)
Valledelcauca	25.71 (1,346.75)
Vaupes	173.93 (1,333.85)
Vichada	3,152.47** (1,334.78)
Observations	360
R ²	0.65
Adjusted R ²	0.60
Residual Std. Error	3,652.89 (df = 320)
F Statistic	14.98*** (df = 39; 320)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01