Econometrics Assignment 4

Chase Bookin & Cole Price

July 14, 2020

1.

A)

```
x <- 36.77097
null <- 0
p <- 0.048
p_div_2 <- p / 2
t <- 1.98
# t is positive because we are testing the coefficient of 36.77 against the
# hypothesis that it is zero.

robust_se <- (x - null) / t
```

To find the t-statistic and Robust Standard Error of the size coefficient, we begin with the p-value provided by Stata. By dividng the p-value by 2 and using a z-score table to match the new CDF value of 0.024, we arrive at the t-statistic of 1.98. We know it is positive because the coefficient for size is positive and being compared to the null hypothesis that the size coefficient is equal to zero. Next, we plug in what we have to the t-stat formula. Dividing the size coefficient of 37.77097 by the t value of 1.98 yields a Robust Standard Error of 18.57. In summary, the t value is found to be 1.98 and the Robust Standard Error is 18.57.

B)

The slope coefficient on size decreased from the first regression to the second regression because we extracted the effect of additional bathrooms on price, and the number of bathrooms in a home is correlated to the size of the home. Essentially, the first regression was overestimating the effect of home size on price through omitted variable bias. Larger homes tend to have more bathrooms, and more bathrooms make homes worth more. By omitting bathrooms in the first regression, its effect was being grouped into the size of the house, yielding the initial overestimate of the first regression and subsequent drop in the size coefficient in the multivariate regression.

2.

B)

Table 1: Data summary

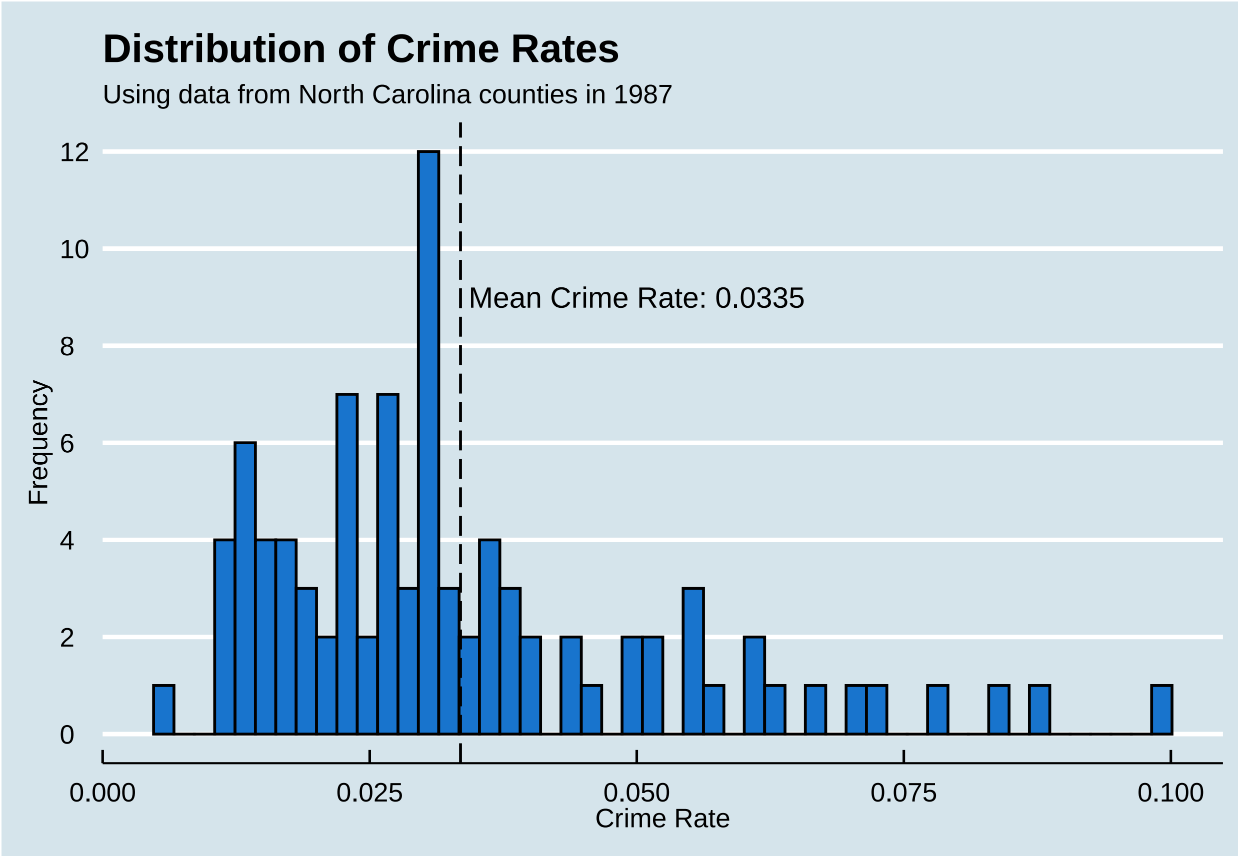
Name	Piped data
Number of rows	90
Number of columns	25
Column type frequency:	
numeric	25
Group variables	None

Variable type: numeric

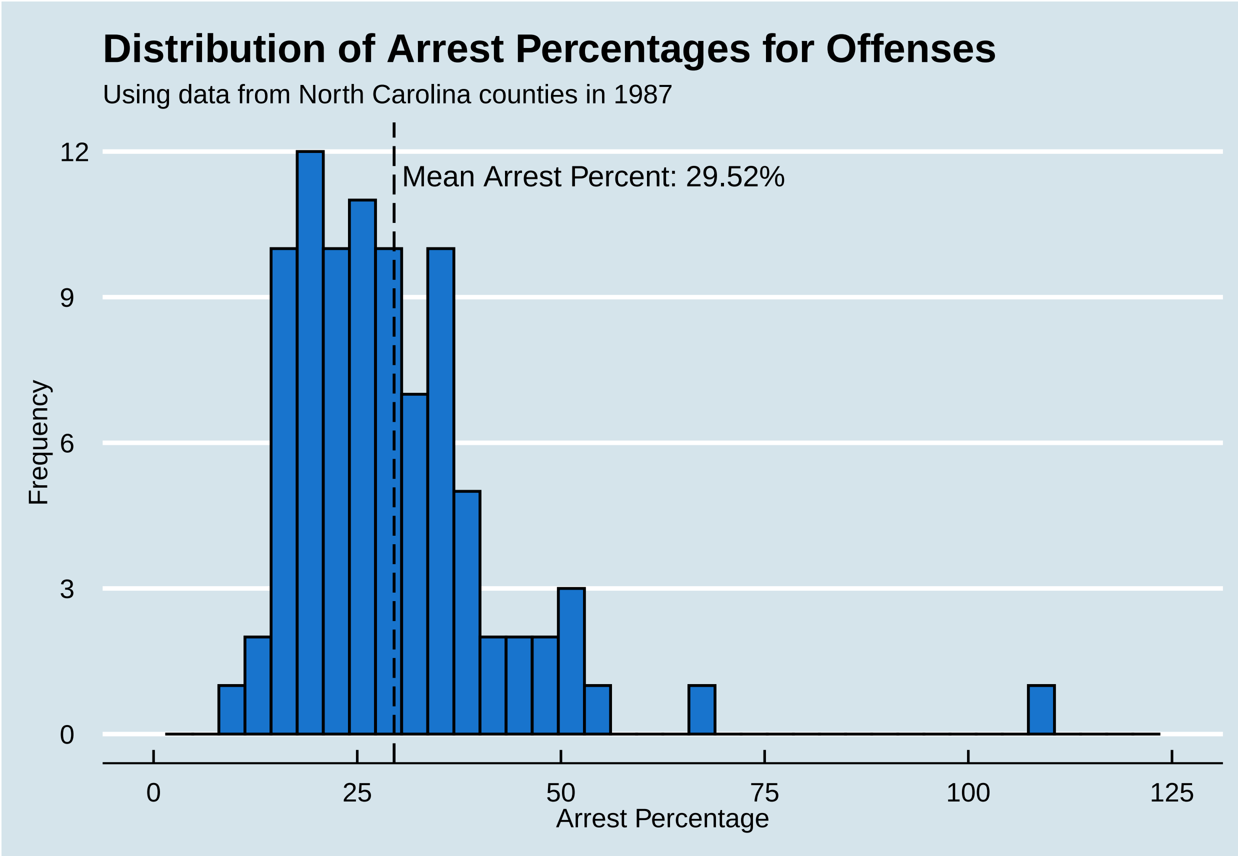
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
county	0	1	100.60	58.32	1.00	51.50	103.00	150.50	197.00
year	0	1	87.00	0.00	87.00	87.00	87.00	87.00	87.00
Crime Rate	0	1	0.03	0.02	0.01	0.02	0.03	0.04	0.10
Arrest Probability	0	1	29.52	13.77	9.28	20.49	27.15	34.49	109.09
Conviction Prob	0	1	55.09	35.42	6.84	34.42	45.17	58.51	212.12
Prison Prob	0	1	41.06	8.07	15.00	36.42	42.22	45.76	60.00
Average Sentence	0	1	9.69	2.83	5.38	7.38	9.11	11.46	20.70
Police per Capita	0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.01
density	0	1	1.44	1.52	0.20	0.55	0.98	1.57	8.83
Tax Revenue per Cap	0	1	38.16	13.11	25.69	30.73	34.92	41.01	119.76
west	0	1	0.23	0.43	0.00	0.00	0.00	0.00	1.00
central	0	1	0.38	0.49	0.00	0.00	0.00	1.00	1.00
east	0	1	0.39	0.49	0.00	0.00	0.00	1.00	1.00
urban	0	1	0.09	0.29	0.00	0.00	0.00	0.00	1.00
Percent Minority	0	1	25.71	16.98	1.28	10.02	24.85	38.18	64.35
Percent Young Male	0	1	0.08	0.02	0.06	0.07	0.08	0.08	0.25
Wage: Construction	0	1	285.35	47.75	193.64	250.75	281.16	314.98	436.77
Wage: Tran & Util	0	1	410.91	77.36	187.62	374.33	404.78	440.68	613.23
Wage: Trade	0	1	210.92	33.87	154.21	190.71	202.99	224.28	354.68
Wage: Finance	0	1	321.62	54.00	170.94	285.56	317.13	342.63	509.47
Wage: Service	0	1	275.34	207.40	133.04	229.34	253.12	277.65	2177.07
Wage: Manufacturing	0	1	336.03	88.23	157.41	288.60	321.05	359.89	646.85
Wage: Federal Gov	0	1	442.62	59.95	326.10	398.78	448.85	478.26	597.95
Wage: State Gov	0	1	357.74	43.29	258.33	329.27	358.40	383.15	499.59
Wage: Local Gov	0	1	312.28	28.13	239.17	297.23	307.65	328.78	388.09

COLE COMMENT ABOVE COLE COMMENT ABOVE COLE COMMENT ABOVE

C)



COLE COMMENT COLE COMMENT COLE COMMENT



COLE COMMENT COLE COMMENT COLE COMMENT

D)

Correlation Matrix

Correlations of crime rate and possible explanatory variables

	crime_rate	arrest_prob	convict_prob	prison_prob	mean_sentence	police_per_cap
crime_rate	1.00	-0.40	-0.39	0.05	0.02	0.17
arrest_prob	-0.40	1.00	-0.06	0.05	0.18	0.43
convict_prob	-0.39	-0.06	1.00	0.01	0.16	0.17
prison_prob	0.05	0.05	0.01	1.00	-0.09	0.05
mean_sentence	0.02	0.18	0.16	-0.09	1.00	0.49
police_per_cap	0.17	0.43	0.17	0.05	0.49	1.00

Data collected from North Carolina Counties in 1987

E)

term	**estimate**	**std.error**	**statistic**	**p.value**
(Intercept)	0.0495	0.00498	9.94	4.89e-16
arrest_prob	-0.000542	0.000128	-4.24	5.5e-05

The regression coefficient of the arrest probability is -0.000542. This means that if arrest probability increases by one percentage point, we expect the crimes committed per person to decrease by 0.000542.

F)

```
estimate <- -0.0005422804
null <- 0
se <- 0.0001278931

t <- (estimate - null) / se
```

From the above calculations as well as the regression output, we find the absolute value of the t-statistic of the regression coefficient for arrest probability to be 4.24. Because the absolute value of the t-statistic is greater than the critical value of 1.96 at a 5% significance level, we reject the null hypothesis that the coefficient for arrest probability is equal to zero.

However, from a practical significance perspective, assuming a population of one million people, an increase of one percent in arrest probability would only drop the total amount of crimes committed by 542 based on the arrest probability coefficient. Most counties in North Carolina do not have one million people, meaning that total crime reduction would be even smaller in these counties.

Additionally, the 50th percentile for crime rate is 0.03. Based on the regression, in order to move from the 50th percentile to the 25th percentile crime rate of 0.02, it is estimated that a county would need to increase its arrest probability by roughly 18.44%, a major increase. This exemplifies the lack of practical significance of the arrest probability coefficient, despite its statistical significance.

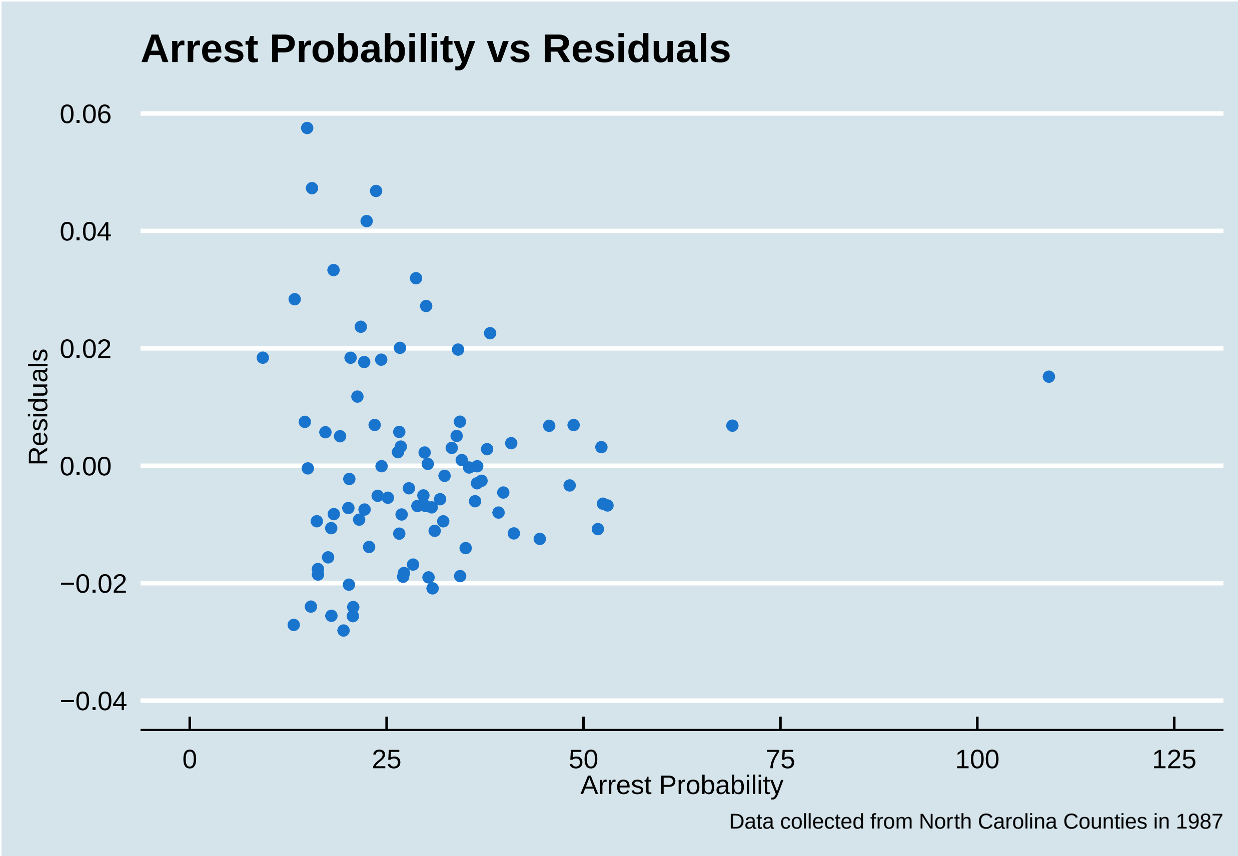
G)

```
intercept <- model1_robust %>%
  filter(term == '(Intercept)') %>%
  pull(estimate)

b1 <- model1_robust %>%
  filter(term == 'arrest_prob') %>%
  pull(estimate)

# y = intercept + b1 * arrest_prob

est_arrest_03 <- (.03 - intercept) / b1 # approx 35.996

est_arrest_02 <- (.02 - intercept) / b1 # approx 54.437

est_arrest_02 - est_arrest_03
```

```
## [1] 18.44064
```

Assuming a county in North Carolina wants to drop from 5 crimes per 100 citizens to 4 crimes per 100 citizens, the estimated increase in arrest probability necessary would be roughly 18.44%. This is the same as the example explained in part F because in each instance the Y variable is being dropped by 0.01 and we are using a linear regression, meaning the slope is the same at all points. Given this, we took the expected arrest probabilities for crime rates of 0.03 and 0.02 and found the difference between them to be 18.44%. This estimated increase in arrest probability required is equal to the change required to drop from 0.05 to 0.04, leaving us with the answer of 18.44%.

H)



Based on this graph, it is clear that the heteroskedasticity-robust estimator for standard error is justified. The residuals have much greater spread for counties with lower arrest probabilities compared to counties with greater arrest probabilities. Because the residuals are not uniformly distributed along the x axis, we need to use the heteroskedasticity-robust standard error.

I)

```
##
## ===========================================================================================================
##                                                    Dependent variable:
##                      -------------------------------------------------------------------------------------
##                                                              crime_rate
##                      coefficient                               OLS
##                         test
##                          (1)              (2)                     (3)                      (4)
## -----------------------------------------------------------------------------------------------------------
## arrest_prob           -0.001***        -0.0003**              -0.0003***              -0.0004***
##                       (0.0001)         (0.0001)               (0.0001)                (0.0001)
##
## density                                0.008***               0.008***
##                                        (0.001)                (0.002)
##
## urban                                                          0.005                   0.037***
##                                                               (0.008)                 (0.005)
##
## Constant              0.050***         0.029***               0.030***                0.042***
##                        (0.005)         (0.004)                (0.004)                 (0.004)
```

```
##
## ----------------------------------------------------------------------------------------------------
## Observations                            90                      90                      90
## R2                                     0.564                   0.566                   0.453
## Adjusted R2                            0.554                   0.551                   0.440
## Residual Std. Error               0.013 (df = 87)         0.013 (df = 86)         0.014 (df = 87)
## F Statistic                  56.255*** (df = 2; 87) 37.368*** (df = 3; 86) 36.001*** (df = 2; 87)
## ====================================================================================================
## Note:                                                             *p<0.1; **p<0.05; ***p<0.01
```