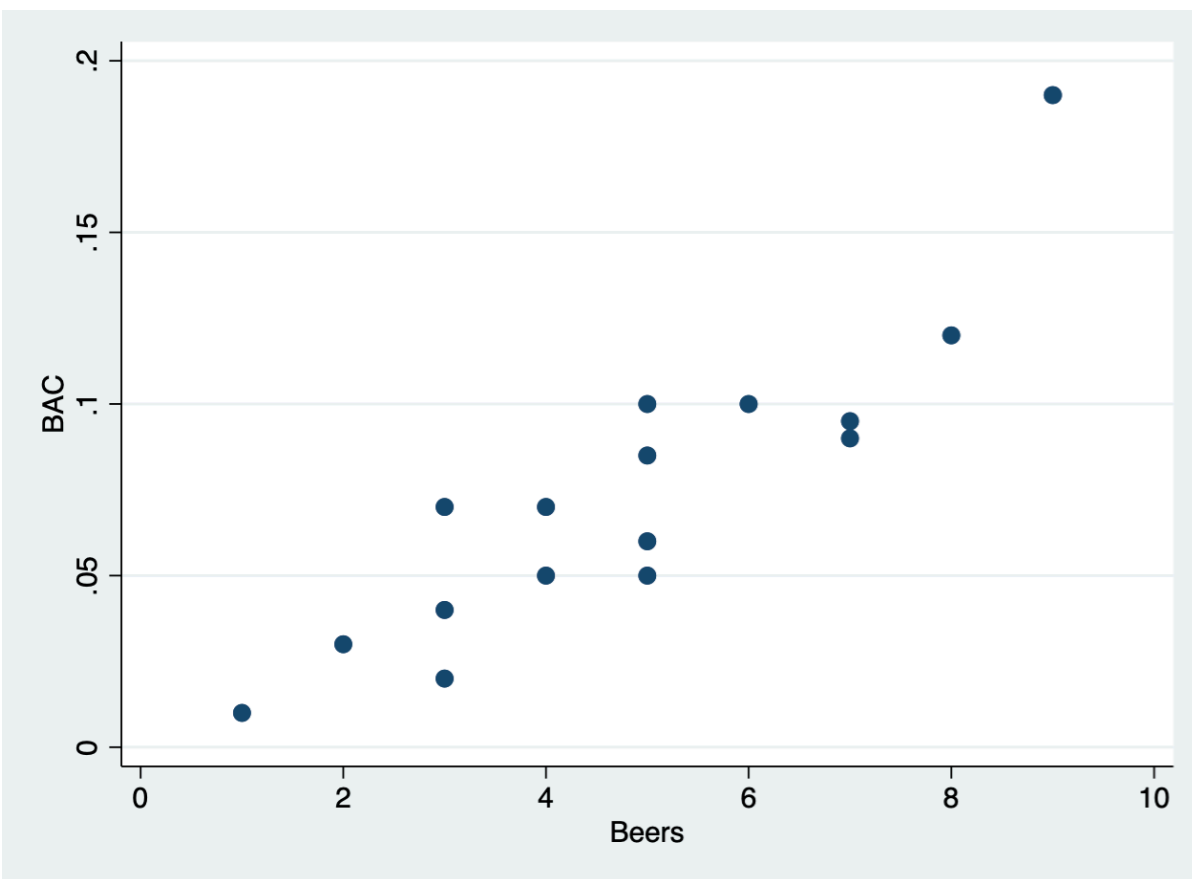


ECO 220: Problem Set 6  
 Lillian Jiang  
 1 December 2018

1. How well does the number of beers a student drinks predict her blood alcohol content?



a)

i) bac and beers are positively related.

ii) The sign of their correlation coefficient is positive.

b)i) All else being equal, an increase of 1 beer leads to 0.018% of increase in one's BAC.

ii) Yes. P-val is nearly 0.00.

iii) Yes. P-val is nearly 0.00.

iv) According to the regression, a student's BAC will be (-0.013) if she has consumed no beers.

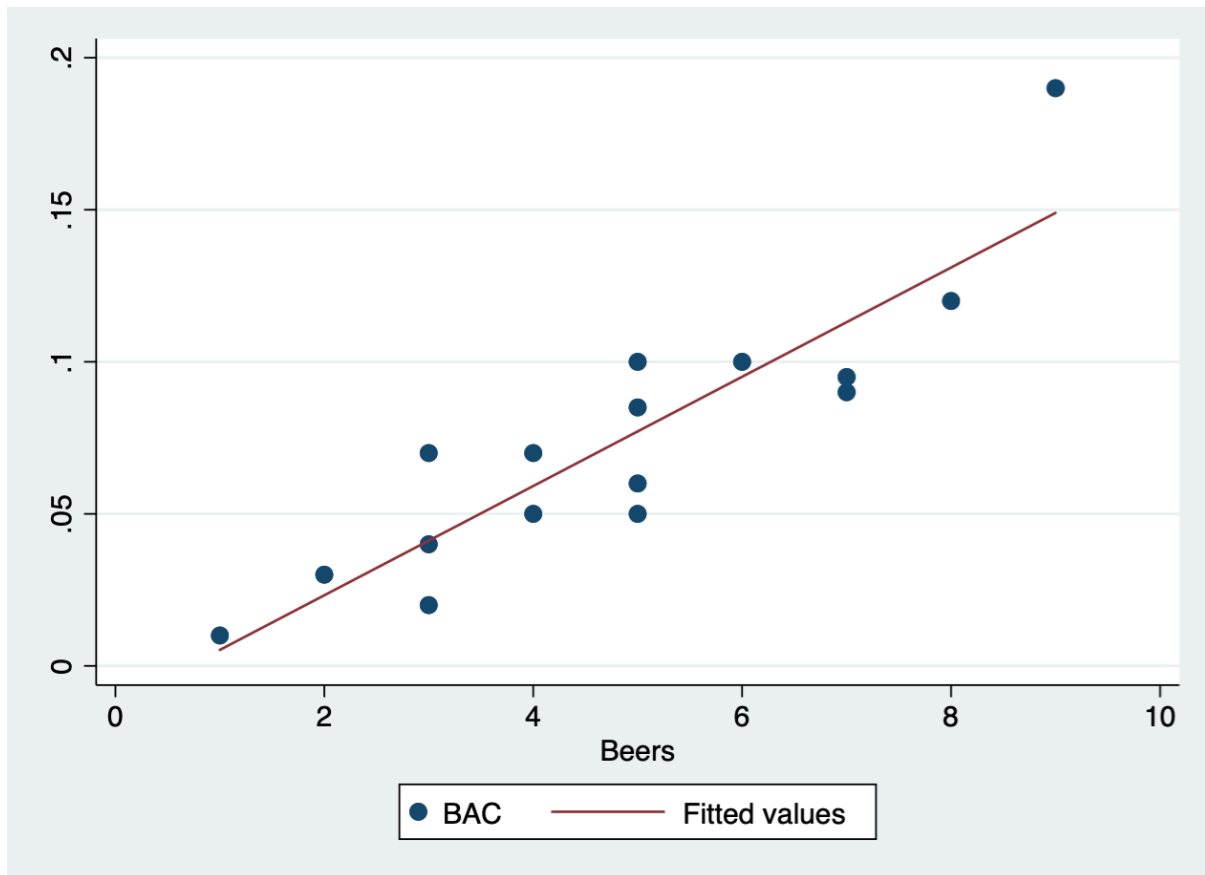
v) I think the slope makes economic of physiological sense, because the more beer one drinks, the higher in blood alcohol content should be; however, the constant or the intercept is not because one's blood alcohol content should not be negative under any circumstance.

c)i)  $\widehat{bac} = -.013 + 0.018 * beers$

d) R-squared = 0.7998.

Beers (the number of beers consumed) explains 79.98% of the total variation in BAC. It is not equal to 1 because the

regressor is not the only thing that explain the outcome. In model, we have an error term that represents everything else but regressor. In regression, we don't have the error term.



e)

f) i)  $H_0: \beta = 0$

$H_a: \beta \neq 0$

ii)  $t\text{-stat} = \frac{\hat{\beta} - 0}{SE(\hat{\beta})} = 7.48$

iii) P-Val = 0.000

iv) We reject  $H_0$ . Beers is statistically significant predictor of BAC at any conventional level.

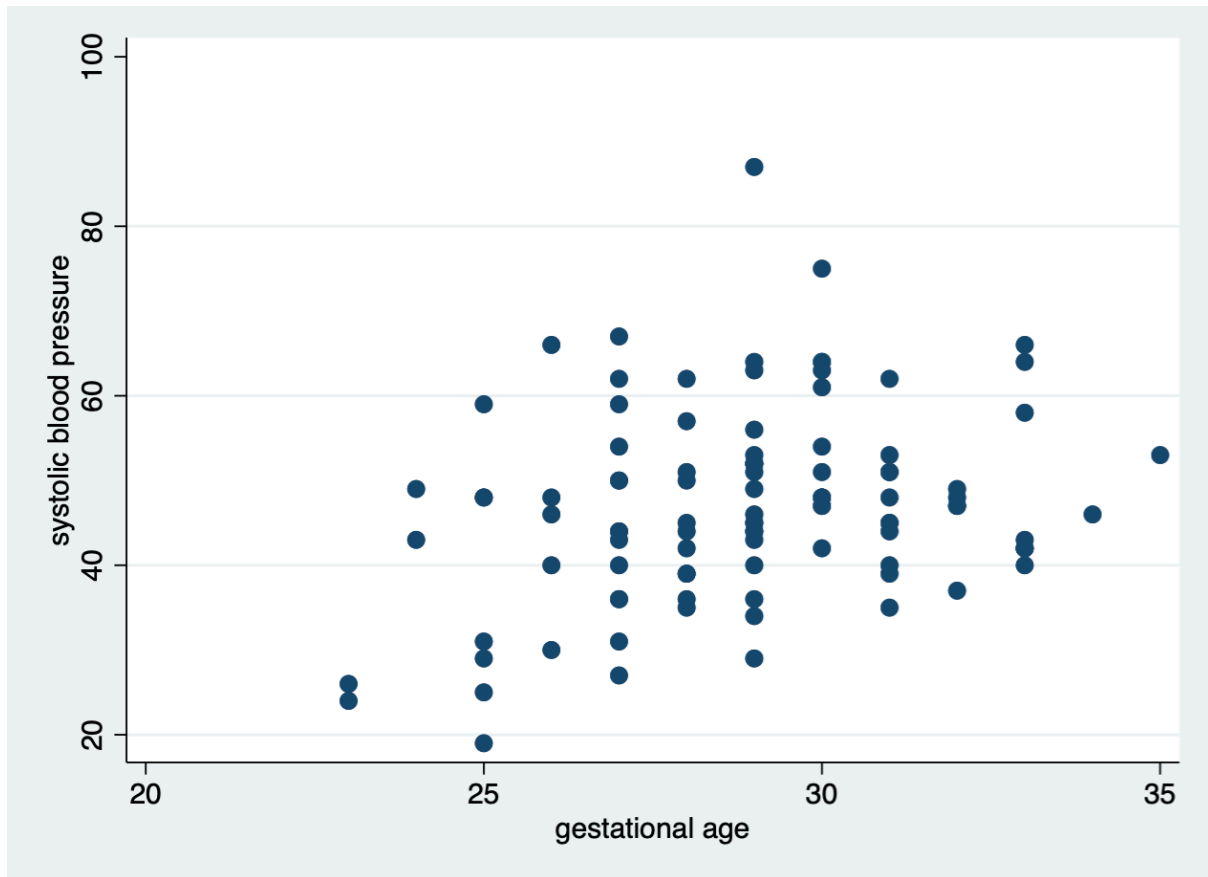
g) BAC for the friend =  $-0.013 + 0.018 \times 6 = 0.095$

h) i) 80% confidence interval: [.265786, .609214]

ii) We are 80% confident that the population proportion of limit will be in .265786 and .609214.

## 2. Infant health and its determinants

a)



They are positively related.

b) i)  $\widehat{sbp} = 25.95 + .0624452 * \text{gestage}$

ii) If gestational age increases 1 week, the systolic blood pressure will increase 0.062 unit of blood.

When there is no gestational age, that is to say, one is not pregnant, her systolic blood pressure will be 25.95.

iii) The slope makes sense because the older gestational age, the higher systolic blood pressure, which means the healthier the kid is. However, the constant or the intercept does not make sense because when the gestage is zero, there should be no systolic blood pressure for infant since the mother does not have the baby yet.

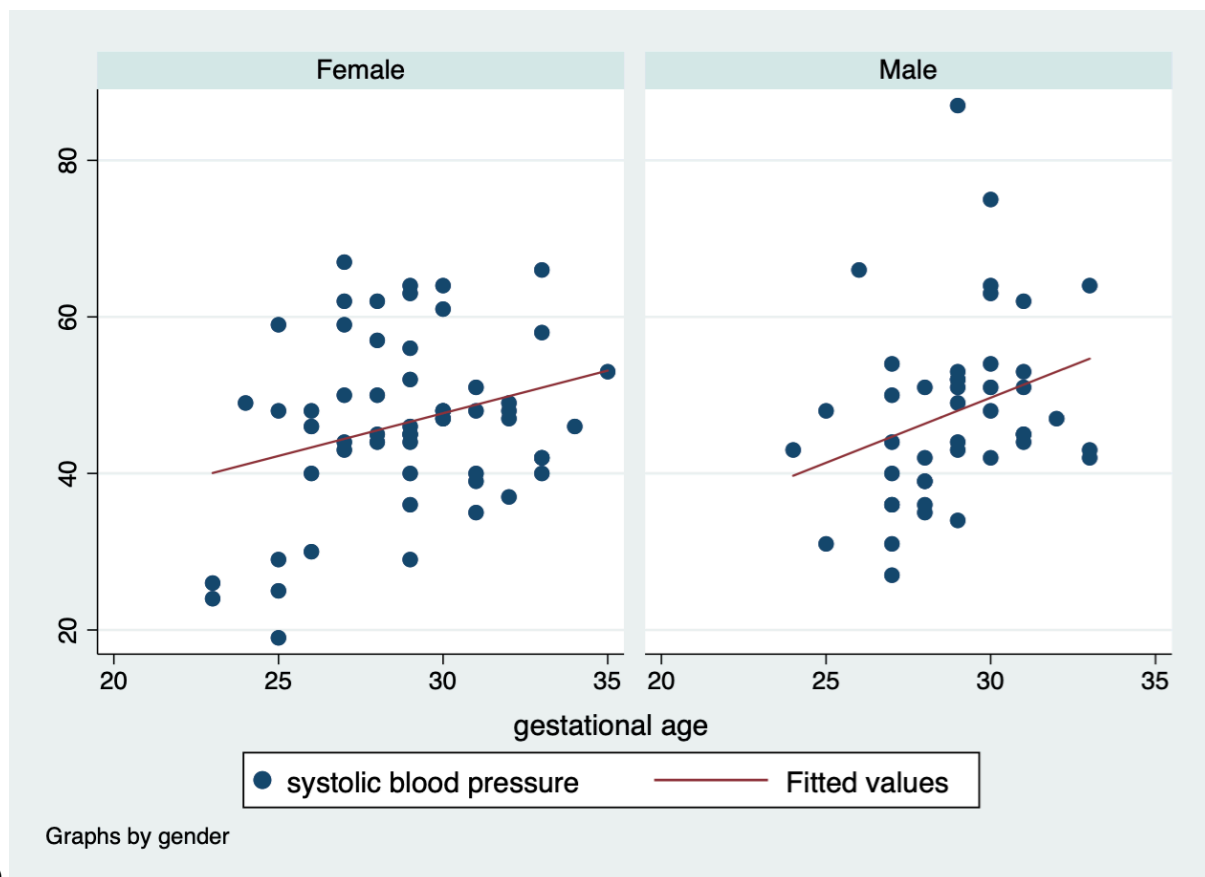
c)  $H_0: \beta = 0$

$H_a: \beta \neq 0$

$$t\text{-stat} = \frac{\widehat{\beta} - 0}{SE(\widehat{\beta})} = 2.90$$

$$P\text{-Val} = 0.005 < 0.05 = \alpha$$

We reject the null hypothesis. Gestage is statistically significant predictor of sbp at any 0.05 level of significance.



d)

The relationship between gestational age and systolic blood pressure is positive for both female and male. However, there is no evidence supporting that it is much stronger for female infants than for male infants. According to the graph, the female's best-fit line's slope is smaller than male's. The  $r^2$  of female infant is also smaller than male's.

### 3. What affects violent crime rates?

a) There are lots of other factors of crime rate. For example, the harshness of the law. If the criminals will receive extreme punishment after being caught, the crime rate would probably decrease.

b) `. sum violent_crime_rate`

Variable	Obs	Mean	Std. Dev.	Min	Max
violent_crime_rate	51	441.6275	241.3983	81	1508

`. sum violent_crime_rate, d`

Percentiles			Smallest	
1%	81	81		
5%	114	110		
10%	237	114	Obs	51
25%	281	167	Sum of Wgt.	51

50%	384		Mean	441.6275
		Largest	Std. Dev.	241.3983
75%	554	787		
90%	707	805	Variance	58273.16
95%	805	812	Skewness	1.757343
99%	1508	1508	Kurtosis	8.609668

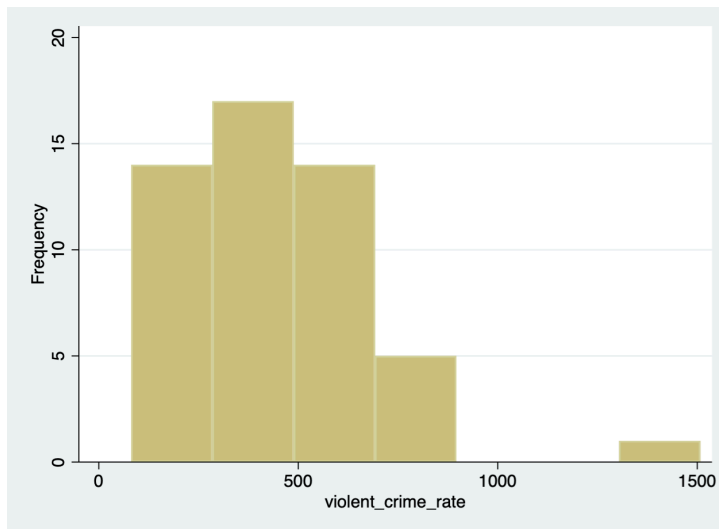
Mean = 441.6275

Median = 384

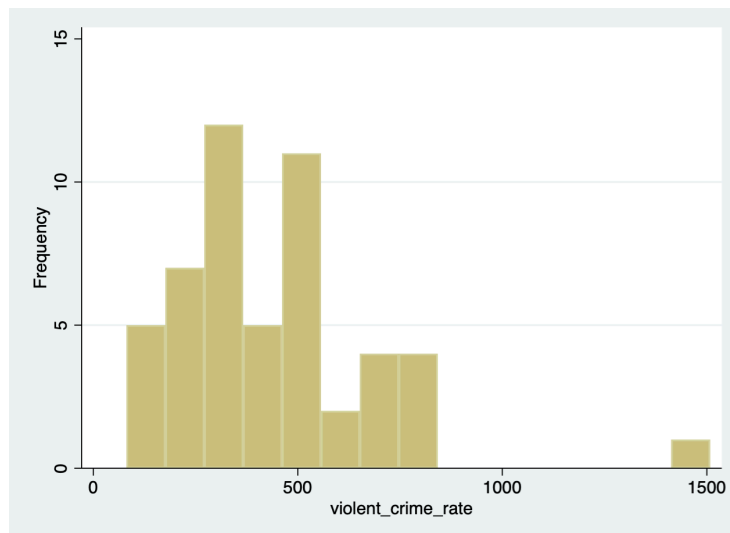
IQR = [75% - 25%] = 554 - 281 = 273

SD = 241.3983

c) Yes, there are large outliers in the data set.



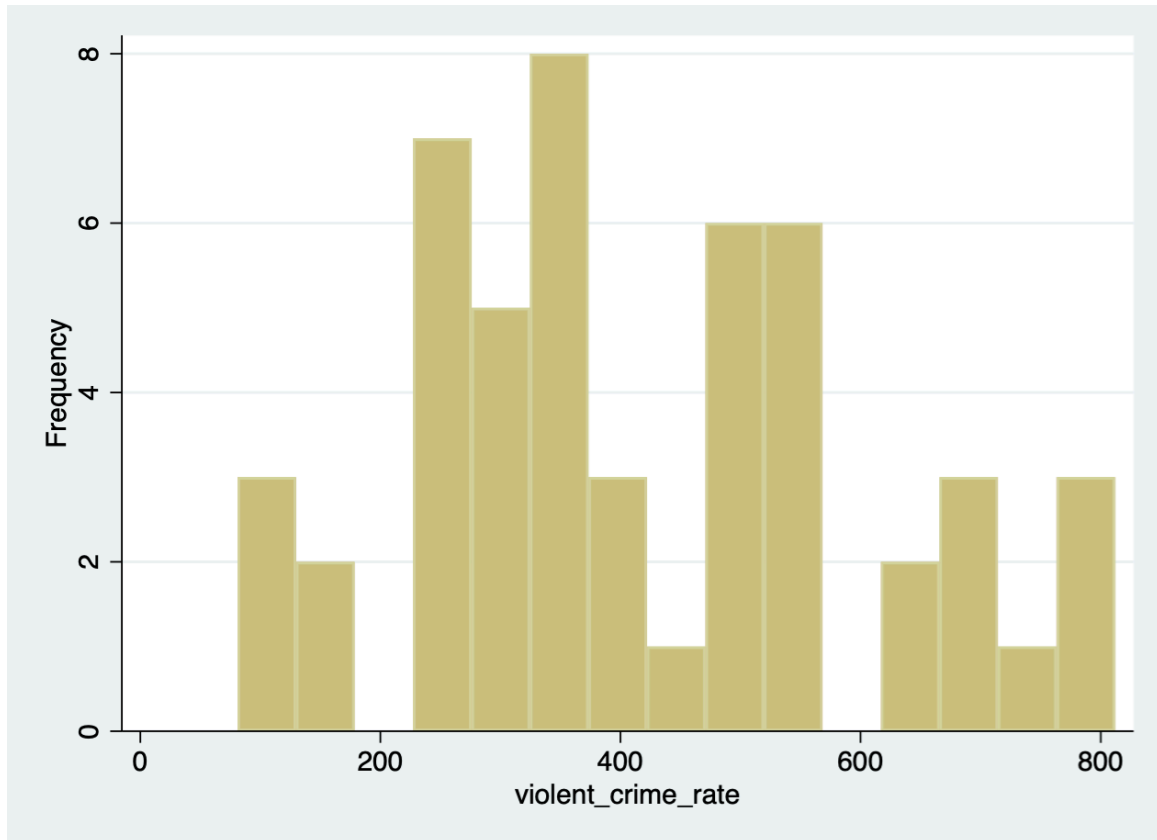
d)



(The first is bin=7; the second one is bin = 15)

i) The bin = 7 has skewness to right and an outlier. The bin = 15 has relatively smaller skewness to right, but still has an outlier.

ii) There are two peaks in the graph.



e)

Yes. It differs. The graph become more symmetric, and have a smaller mean.

f) The median is 190. The middle observation of dollar expenditure per capita on police in this sample is 190 millions.

g) The mean for  $\text{police} < 190 = 363$

The mean for  $\text{police} \geq 190 = 517.2308$

I think the difference come from the reason that there is association between dollar expenditure and violent crime rate. Because the violent crime rate is high, so the state put more money on police in order to increase safety.

h) i) 0.1679 of violent crime rate is explained by the regression line with in poverty.

0.5939 of violent crime rate is explained by the regression line with in single\_parent.

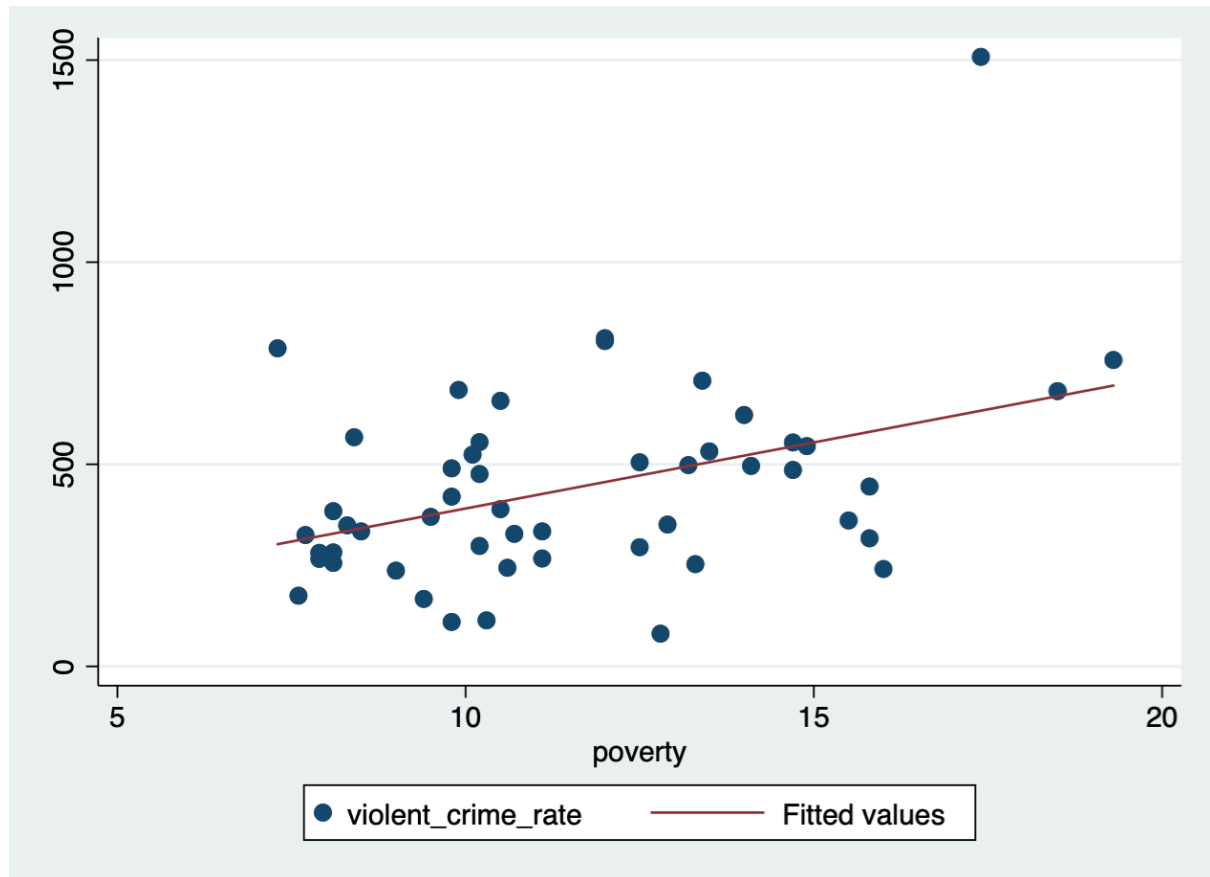
0.1799 of violent crime rate is explained by the regression line with in single\_parent.

ii) Poverty is positively associated with violent crime rate, and the association is statistically significant.

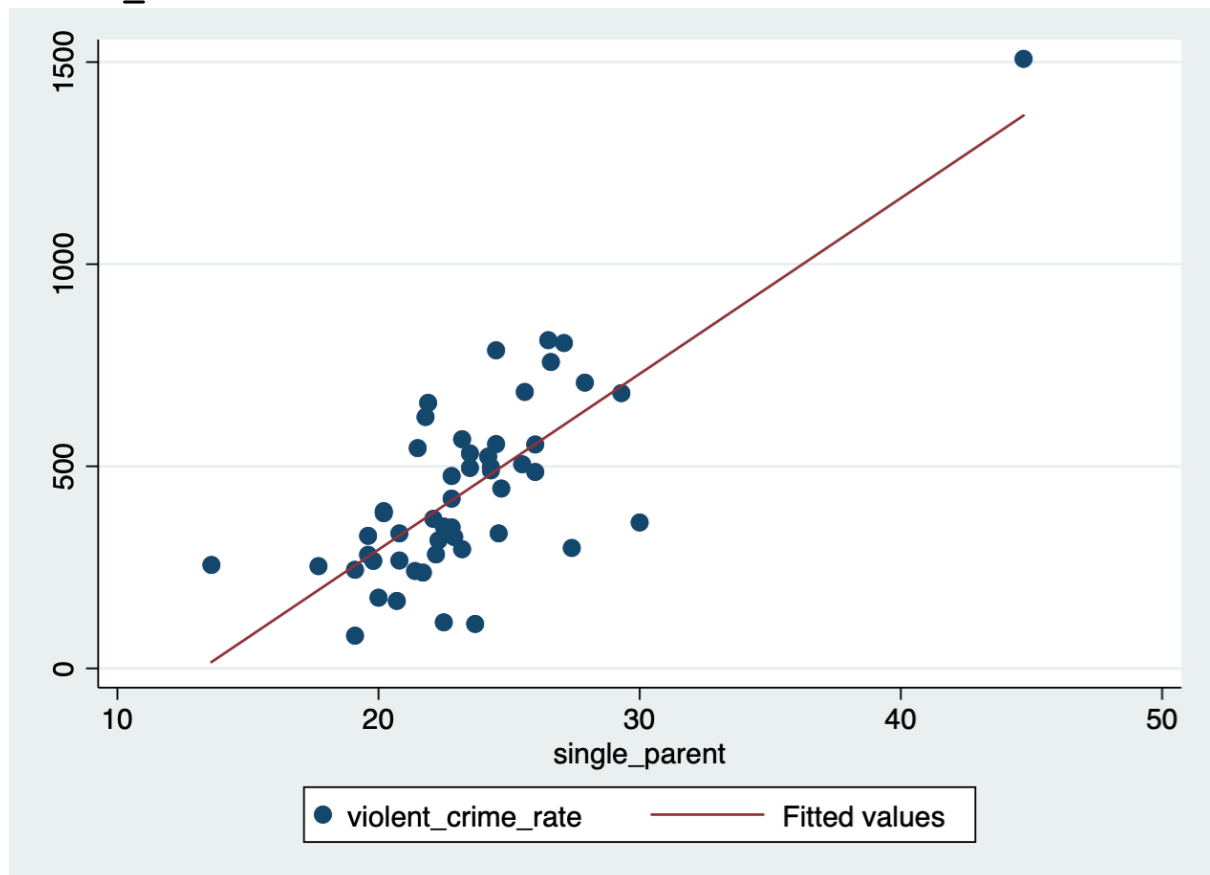
Single\_parent is positively associated with violent crime rate, and the association is statistically significant.

Unemployed is positively associated with violent crime rate, and the association is statistically significant.

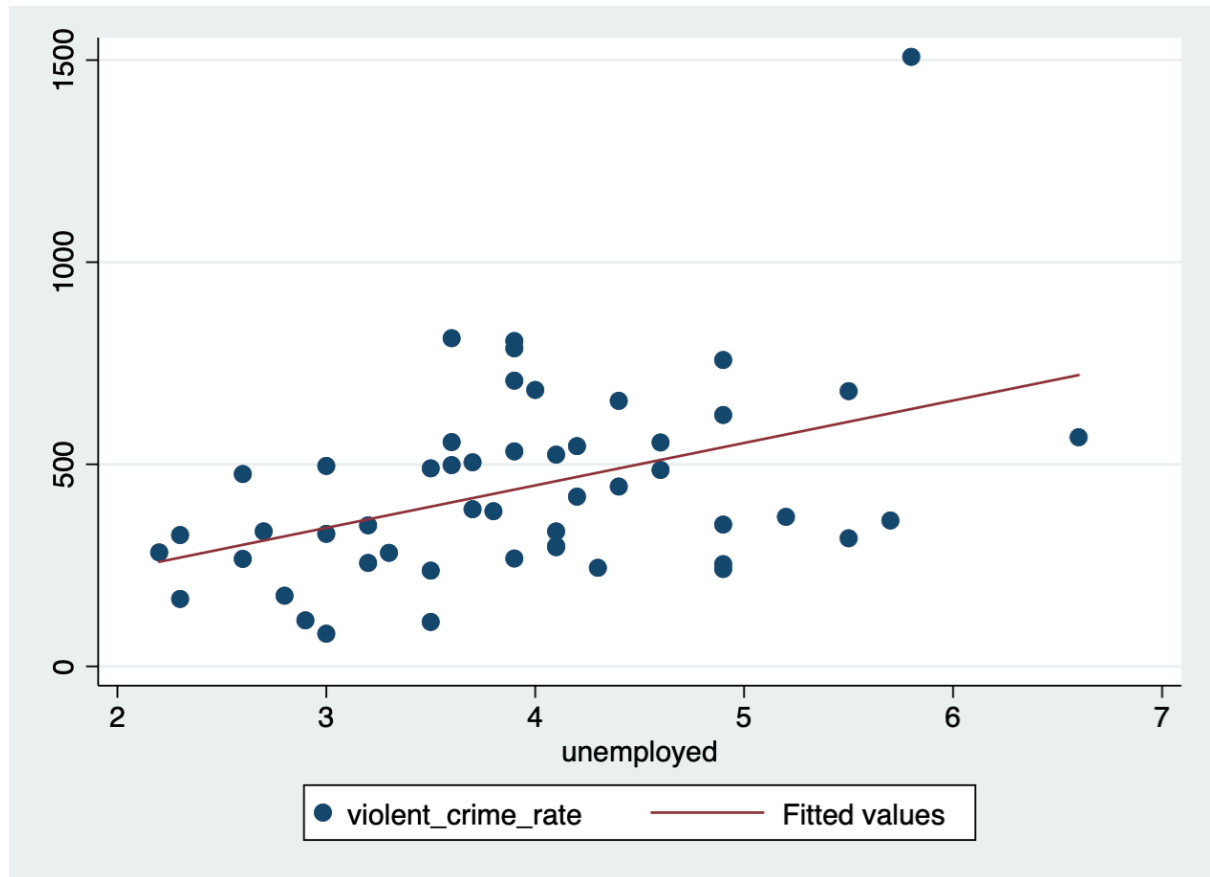
iii)poverty



single\_parent



unemployed



iv) `Single_parent`. Because it has the highest R-squared value.

ii) The mean violent crime rate for states that considered `literate` is 357.9063

The mean violent crime rate for states that considered `illiterate` is 582.6316

iv) The slope = -224.7253

The regression model is

$$\widehat{\text{violent crime rate}} = 582.6316 + (-224.7253) * \text{literate}$$

When `literate == 1`,  $\widehat{vcr1} = 582.6316 - 224.7253$

When `literate == 0`,  $\widehat{vcr0} = 582.6316$

$$\widehat{vcr1} - \widehat{vcr0} = -224.7253 = \text{regression coefficient}$$

j) 60.52% of variability in the violent crime rate is explained by the regression line with in the regressors, including `poverty`, `single_parent`, and `unemployed`.

Only `poverty` change sign from negative to positive.

`Single_parent` is statistically significant predictor and it is the only one in the multiple regression.