

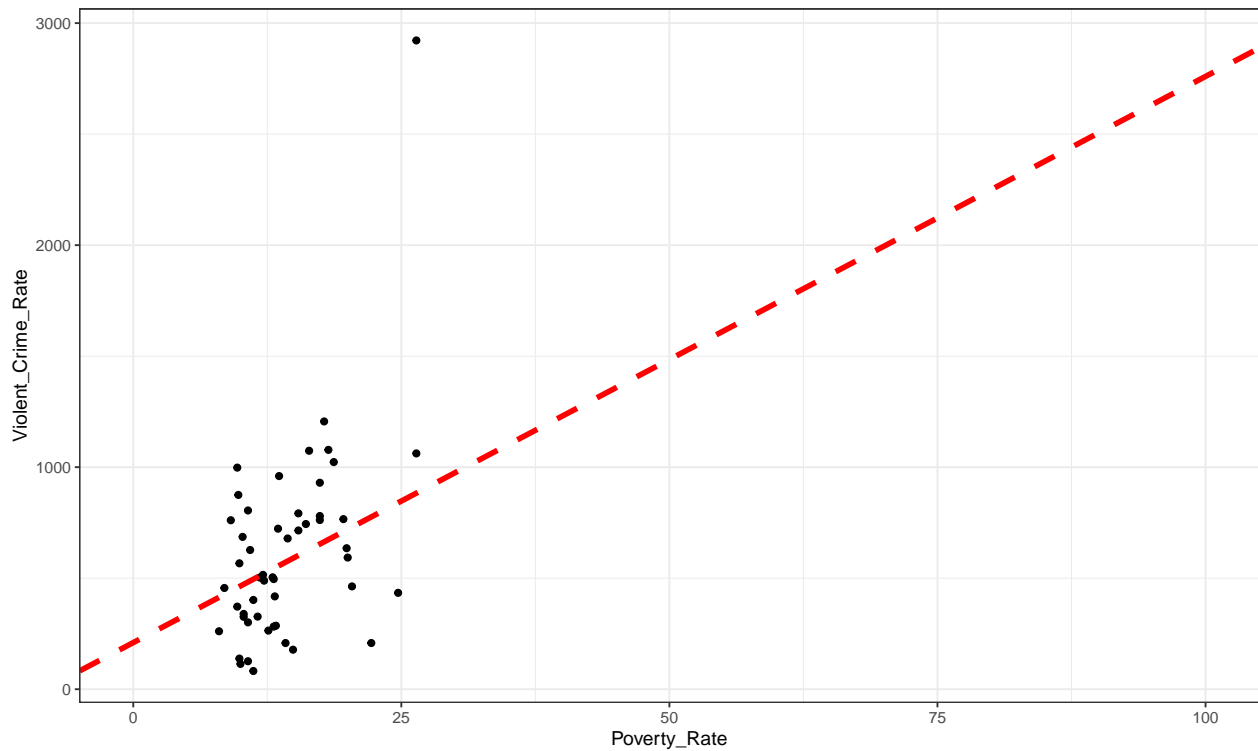
HW 1

Adon Rosen

Date: 2020-02-11

Question 3

a.



b.

The y intercept here indicates what a predicted crime rate would be given a poverty rate of 0. Specifically, if there was no poverty rate, then we would expect a violent crime rate of 209.9.

c.

Predicted value = $209.9 + 25.5 * 10.7$; this is equal to 482.75

d.

The residual is equal to $Y - \hat{Y}$; here $Y = 805$ and $\hat{Y} = 482.75$. So the residual is equal to 322.25. This residual indicates we underpredicted the observed value for the state of Massachusetts.

e.

If two states differ by 10 in their poverty rates they will differ by 255 in their predicted violent crime rate.

f.

First find the minimum value: 413.9

Now find the maximum value: 839.75

Now find the differences to obtain the range of predicted values: 425.85

g.

The pearson correlation between poverty rate and violent crime rate is positive. As poverty rate increases, so does violent crime.

Question 4

a.

This equation is not realistic, because the data would quickly reach the ceiling of possible college GPA values. For instance a student with a high school GPA of .5 would be predicted to have a college GPA of 4.0, and anyone with a GPA above .5 would be expected to have a GPA greater than 4.0, which is not possible on the provided scale.

b.

The slope here indicates that a 1 unit increase in high school GPA yields a .7 increase in an estimated college GPA.

c.

1. $\hat{Y}_3 = 2.6$
2. $\hat{Y}_4 = 3.3$

d.

Here the y-intercept is = 0; and the slope is 1. This indicates that there is an identical relationship in high school GPA and predicted college GPA.

Question 10

a.

$$\hat{Y} = -37.229 + 42.969\beta_{bedrooms}$$

The slope here is 42.969 indicating for every additional bedroom the houses estimated price will increase by 42.969 thousands. This is a positive relationship such that as one increases so does the other.

b.

1. $\hat{Y}_2 = 48.709$
2. $\hat{Y}_3 = 91.678$
3. $\hat{Y}_4 = 134.647$

c.

The $\hat{Y}_3 = 91.678$ whereas the observed value is 48.5. The residual is equal to $Y - \hat{Y}$ which equals to -43.178

d.

In order to calculate the correlation from the slope first multiply the slope of the regression line by the standard deviation of X and then divide by the standard deviation of Y.

Slope: 42.969

SDx: 0.6070

SDy: 44.1841

Formula: $(\text{Slope} * \text{SDx}) / \text{SDy} = 0.590307$

e.

The coefficient of determination is: 0.3484 ; this means that our model can explain roughly 34% of the variation observed in the outcome

f.

The standard error for the β_{bedrooms} is equal to 6.160 - this value reflects the distribution of the sampling statistic.

g.

Here the p value for $\beta_{\text{bedrooms}} = 0.0001$, so we reject the null hypothesis that $\beta_{\text{bedrooms}} = 0$ and conclude that there is a non zero relationship between bedrooms and predicted house cost.

h.

The 95% confidence interval for the coefficient is equal to 42.969 ± 12.0736 . This value indicates if we were to repeat this sampling procedure an infinite number of times we would expect 95% of the obtained coefficients to fall within this range.

i.

$$\hat{Y}_{4-upper}=182.9414$$

$$\hat{Y}_{4-lower}=86.3526$$

$$\hat{Y}_{2-upper}=72.8562$$

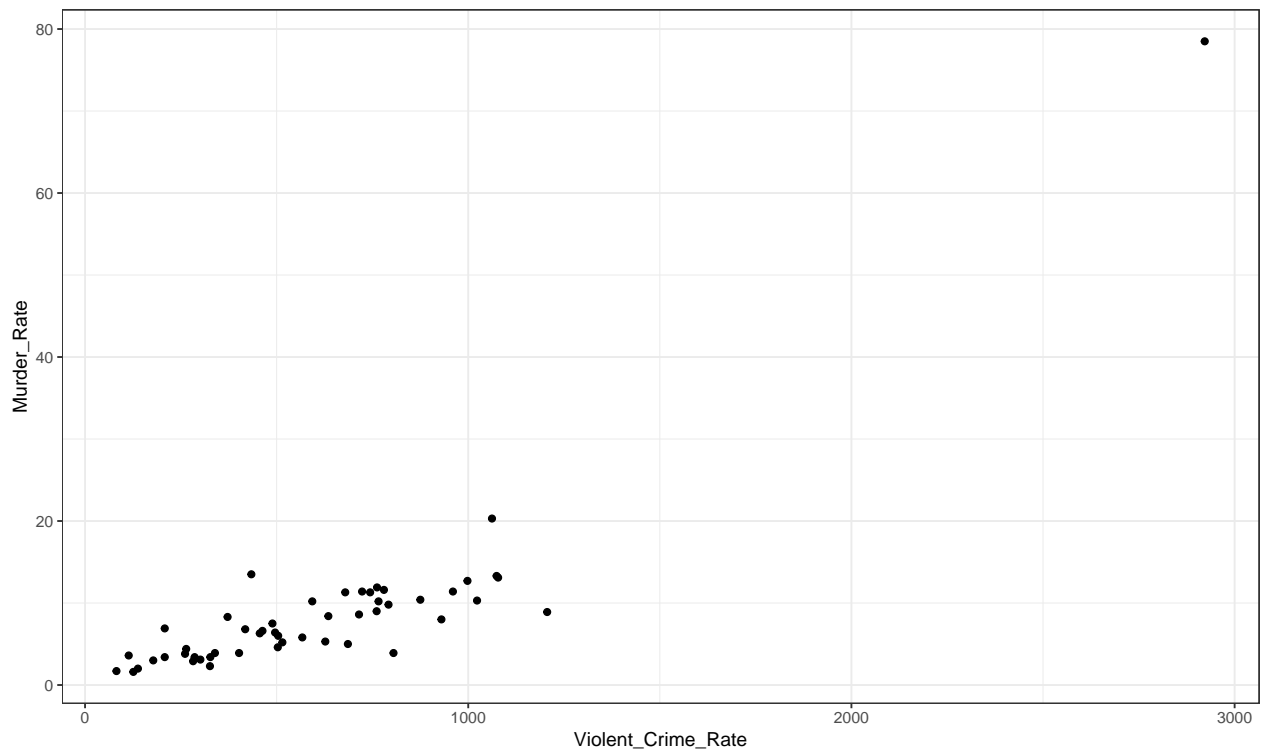
$$\hat{Y}_{2-lower}=24.5618$$

j.

The Root MSE is the standard deviation of the residuals. So on average we expect the residual to be 35.86120 units away from the predicted value

Question 12

a.



There appears to be a pos relationship between these two variables

b.

Call:

```
lm(formula = Murder_Rate ~ Violent_Crime_Rate, data = table9_1)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-12.5994	-2.5561	0.1884	2.0846	20.0512

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.468432	1.210230	-3.692	0.000559 ***
Violent_Crime_Rate	0.021532	0.001608	13.389	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.016 on 49 degrees of freedom
Multiple R-squared: 0.7853, Adjusted R-squared: 0.781
F-statistic: 179.3 on 1 and 49 DF, p-value: < 2.2e-16

c.

The $\hat{Y}_{DC} = 58.4487628$

This means we predict there to be about 58.4487628 murder rate for DC

d.

The $\hat{Y}_{612.8} = 8.7265221$

e.

In order to calculate the correlation from the slope first multiply the slope of the regression line by the standard deviation of X and then divide by the standard deviation of Y.

Slope: 0.0215322

SDx: 441.1

SDy: 10.7

Formula: (Slope * SDx) / SDy = 0.8876514

f.

Question 35

The answer is B - A correlation of .3 shares has a coefficient of determination of .09

Question 37

The answer is G - Both a coefficient and correlation are affected by outliers

Question 38

Linear regression has a set of assumptions which include:

1. There is a linear relationship between Y and X
2. The error are homoscedastic
3. The error are normally distributed
4. The sample observations are selected randomly and independent of each other