Gabriella Weis
Mr. Lewis
Honors Applied Statistics
9 September 2023

Assignment #5, Introduction to R

1.) Acquired data set USArrests.

2.) Created a plot comparing assault rates in each state as a predictor to murder rates.

Variables:

x = Assault rate
y = Murder rate

Code:

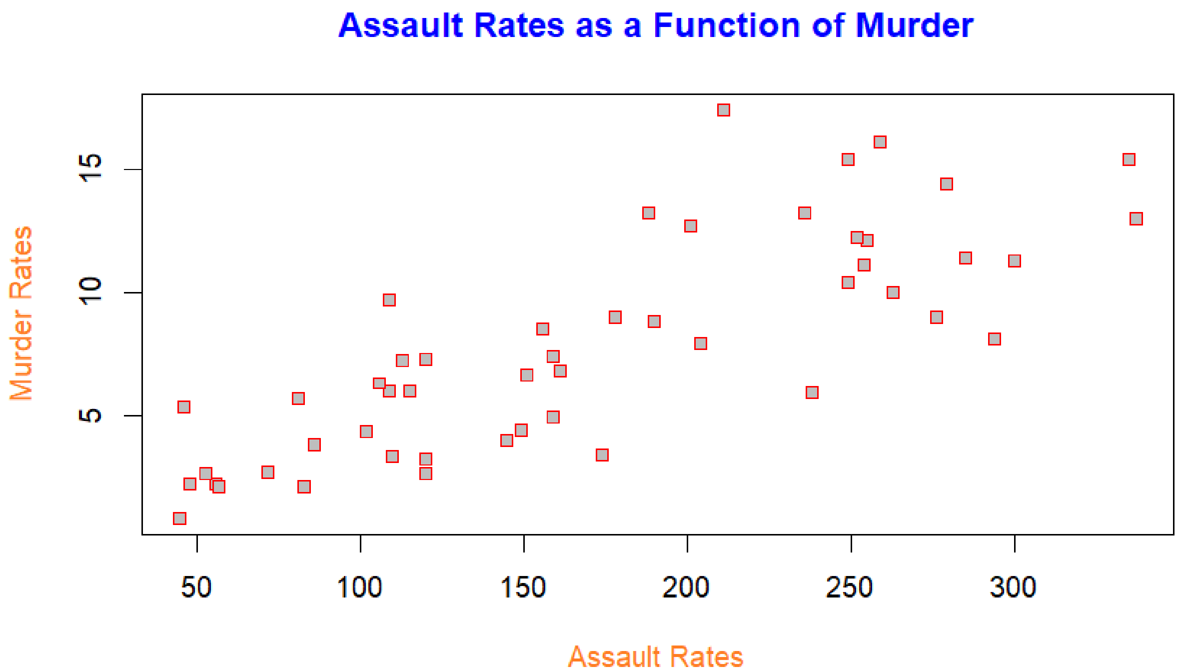
```
setwd("C:/Users/weisgab000/Documents/Applied Stats/Intro to R/Assignment #5")

data("USArrests")

attach(USArrests)

plot(Assault,Murder,main="Assault Rates as a Function of Murder",xlab="Assault Rates",ylab="Murder Rates",pch=22,col="red",bg="grey",col.main="blue",col.lab="chocolate1")
```

Scatter Plot:



Interpretation:

There appears to be a moderate, positive, linear relationship between Assault and Murder rates.

3.) Ran a regression analysis representing the relationship of assault rates predicting murder rates.

Code and summary output:

lm(Murder~Assault)

```
Coefficients:
(Intercept)     Assault
    0.63168     0.04191
```

regres1<-lm(Murder~Assault)

summary(regres1)

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.8528 -1.7456 -0.3979  1.3044  7.9256

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

(Intercept) 0.631683   0.854776   0.739   0.464
Assault   0.041909   0.004507   9.298 2.6e-12 ***

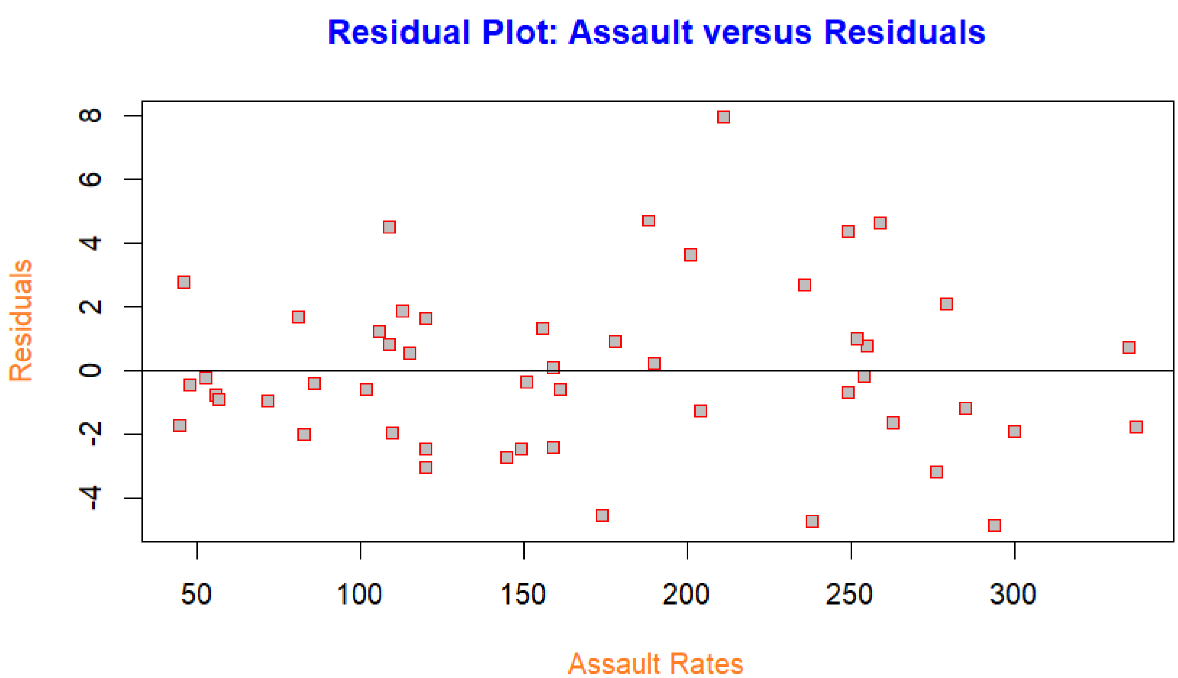
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.629 on 48 degrees of freedom
Multiple R-squared: 0.643,	Adjusted R-squared: 0.6356
F-statistic: 86.45 on 1 and 48 DF, p-value: 2.596e-12

abline(regres1)

4.) Checked the conditions needed in order to run linear regression inferences with a residual plot.

Residual plot:



Code:

plot(Assault,resid(regres1),main="Residual Plot: Assault versus Residuals",xlab="Assault Rates",ylab="Residuals",
pch=22,col="red",bg="grey",col.main="blue",col.lab="chocolate1")

abline(0,0)

Interpretation:

The residual plot appears to be randomly distributed, signifying that the residuals have equal variance and are independent, and that a linear model fits the data well.

5.) Predicted murder rate for a state having an assault rate of 300.

Variables:

$\hat{y}$ = predicted Murder rate
x = Assault rate

Formula, substitute, answer:

$\hat{y} = a + bx$

$\hat{y} = 0.63168 + 0.04191x$

$\hat{y} = 0.63168 + 0.04191(300)$

$\hat{y} = 13.20468$

Interpretation:

The predicted Murder rate of a state with an Assault rate of 300 is 13.20468.

6.) A 95% confidence interval for the slope, interpreted.

Formula, substitute, answer:

b +/- $t^*(S_b)$

0.04191 +/- (-2.0106)(0.004507)

(0.03284621, 0.05097104)

Interpretation:

We are 95% confident that the true slope (β), the unit increase in Murder rate (y) for a one unit increase in Assault rate (x), lies within the interval (0.03284621, 0.05097104). We used a method to calculate this interval that will capture β 95% of the time in the long run.

7.) Found and interpreted S_e.

Formula, substitute, answer:

$$S_e = \sqrt{\frac{SSResid}{n-2}} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$$

$$S_e = \sqrt{\frac{331.85}{50-2}}$$

$$S_e = 2.629$$

Interpretation:

The standard error of the regression line depicting Murder rate (y) as a function of Assault rate (x) is 2.629. This means that the average residual from the regression line is 2.629 units (murders).

8.) Finding evidence for a statistically significant relationship between assaults and murders.

Intro:

Model Utility Test
x = Assault rate
y = Murder rate
β = true slope of Assault rates (x) vs. Murder rates (y)
b = predicted slope of Assault rates (x) vs. Murder rates (y)

$H_o: \beta = 0$
$H_a: \beta \neq 0$

$\alpha = 0.05$

Conditions:

☑ Linear: data appears linear in scatter plot (see question #2)
☑ Indepence: assume independence
☑ Normal: data appears to be normally distributed in residual plot (see question #4)
☑ Equal variance: points appear equally varied from y=0 line on residual plot (see question #4)

Calculations:

$$t = \frac{b-\beta}{S_b}$$

$t = \frac{0.04191-0}{0.004507}$

$t = 9.2989$ w/df = 48

p-value = $p(t > 9.2989)x2 = 2.6x10^{-12} \approx 0$

Interpretation:

Since the p-value (0) < α(0.05), we reject the null hypothesis. Therefore, there is enough evidence to suggest a useful linear relationship between Assault and Murder rates.