

Statistics 12, Lab 3

Michael Wu
UID: 404751542

February 19th, 2019

Exercise 1

a) I ran the following code.

```
linear_model <- lm(soil$lead ~ soil$zinc)
summary(linear_model)
```

This output the following statistics.

Call:

```
lm(formula = soil$lead ~ soil$zinc)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|---------|---------|--------|--------|---------|
| | -79.853 | -12.945 | -1.646 | 15.339 | 104.200 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | 17.367688 | 4.344268 | 3.998 | 9.92e-05 *** |
| soil\$zinc | 0.289523 | 0.007296 | 39.681 | < 2e-16 *** |

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

Residual standard error: 33.24 on 153 degrees of freedom

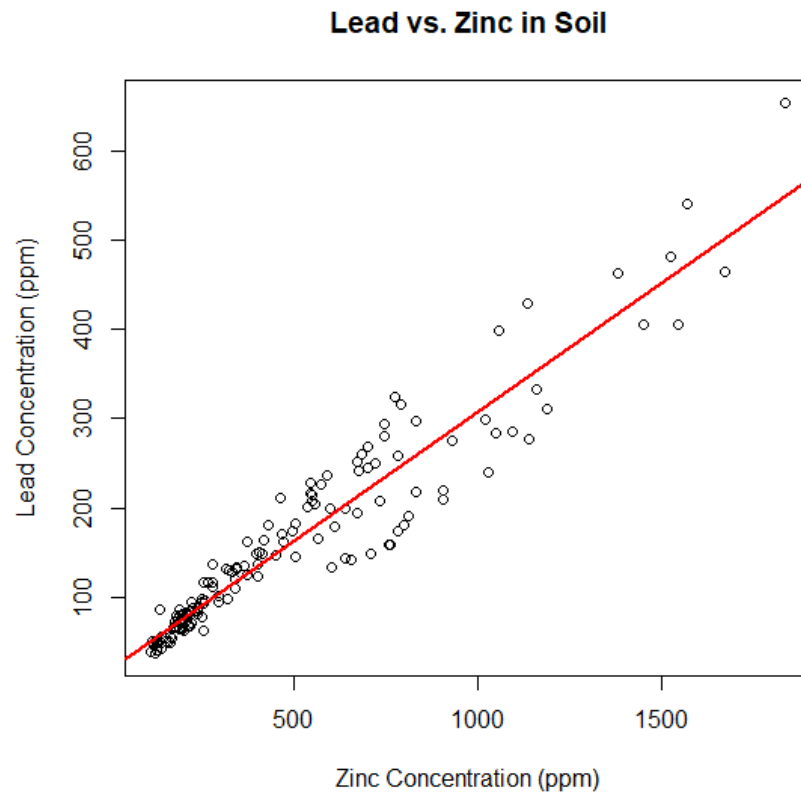
Multiple R-squared: 0.9114, Adjusted R-squared: 0.9109

F-statistic: 1575 on 1 and 153 DF, p-value: < 2.2e-16

b) I ran the following code.

```
plot(soil$lead ~ soil$zinc, xlab="Zinc Concentration (ppm)",  
     ylab="Lead Concentration (ppm)", main="Lead vs. Zinc in Soil")  
abline(linear_model, col="red", lwd=2)
```

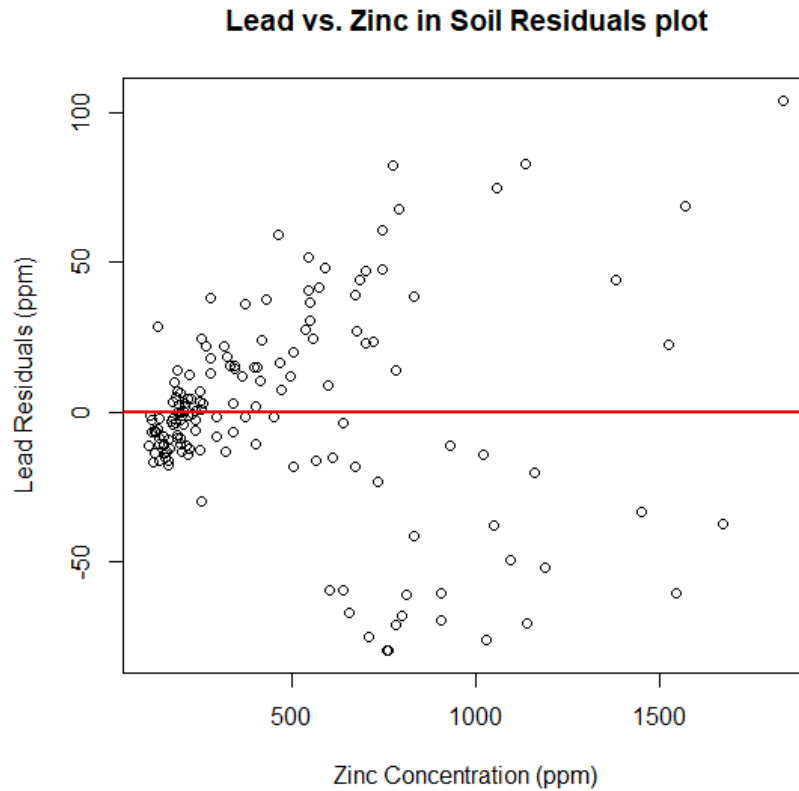
This generated the following plot.



c) I ran the following code

```
plot(linear_model$residuals ~ soil$zinc, xlab="Zinc Concentration (ppm)",  
     ylab="Lead Residuals (ppm)", main = "Lead vs. Zinc in Soil Residuals plot")  
abline(a=0, b=0, col="red", lwd=2)
```

This generated the following plot.



d) The equation of the regression line is the following.

$$y = 0.289523x + 17.367688$$

e) We would expect the lead concentration to be 306.89 ppm.

f) We would expect the lead concentration to be 28.95 ppm higher.

g) The R-squared value is 0.9114. In this context it means that 91.14% of the variation in the lead concentration can be determined by the variation in zinc concentration.

h) I believe that linearity, symmetry, and are met for this data. The data appears very straight and the residuals look symmetrical. However, the condition for equal variance is not met. The residuals appear to become more spread out as the zinc concentration becomes greater. This would indicate that our model does not capture all the details of our system. Perhaps of the logarithms of both zinc and lead would be more appropriate to compare with each other.

Exercise 2

a)

b)

c)

Exercise 3

a)

b)

c)

d)

e)

f)

Exercise 4

a)

b)

c)

d)

e)