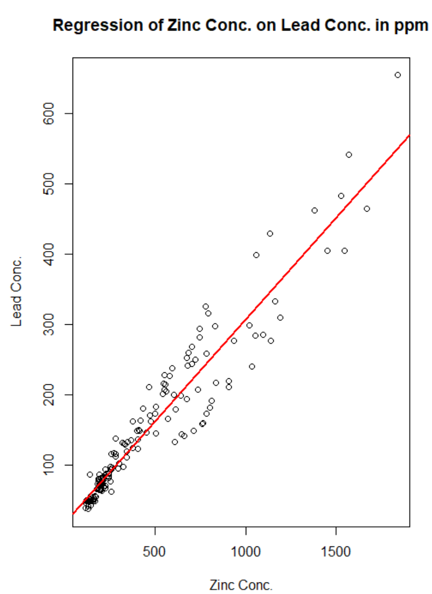
Stats 12 Lab Report 3

Brian Tehrani

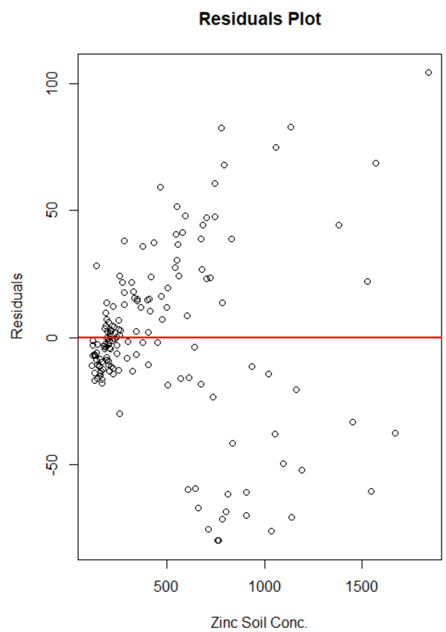
UID: 604715464

1. Exercise 1

   a.
```
> #a
> Lead_Soil_Reg <- lm(soil$lead ~ soil$zinc)
> summary(Lead_Soil_Reg)

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.

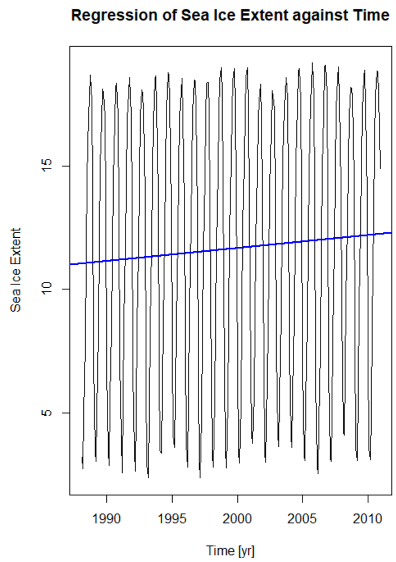


c.
d. y =a + bx where a = 17.367688, b = 0.289523
e. Using the regression line from d): if Zinc = 1000 ppm, Pb = 306.890688 ppm
f. We would heve to use the slope: 100 * 0.289523 = 28.9523. Therefore, lead conc. at A would be 28.9528 times higher than B
g. The R-squared value is 0.9114. This value means that 91.14% of the variation in Lead conc. is explained by the Zinc conc.
h. The linearity of the linear regression seems to be met as the data points seem to distribute evenly about the line, although the data around Zinc concentrations of 550 – 1250 seems to deviate away from the line. The residuals plot shows the data points to be distribute close to the line before the 500 Zinc concentration mark, but then deviate greatly away from the line indicating that the variances are not equal.

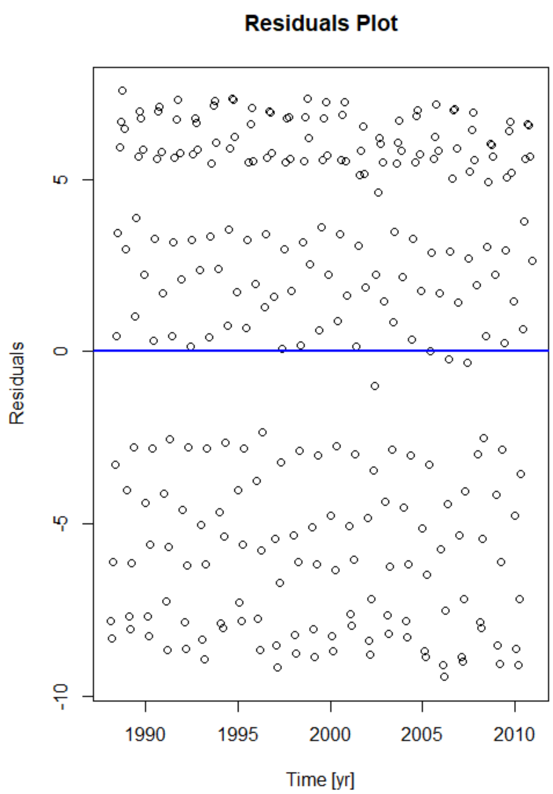
2. Exercise 2

a.
```
> #a
> SeaExt_Time_Reg = lm(ice$Extent ~ ice$Date)
> summary(SeaExt_Time_Reg)

Call:
lm(formula = ice$Extent ~ ice$Date)

Residuals:
   Min     1Q Median     3Q    Max
-9.445 -5.439  1.442  5.599  7.564

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.011e+01  1.558e+00   6.486 4.11e-10 ***
ice$Date    1.438e-04  1.411e-04   1.019    0.309
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 5.654 on 273 degrees of freedom
Multiple R-squared: 0.003787, Adjusted R-squared: 0.0001377
F-statistic: 1.038 on 1 and 273 DF, p-value: 0.3093



b.



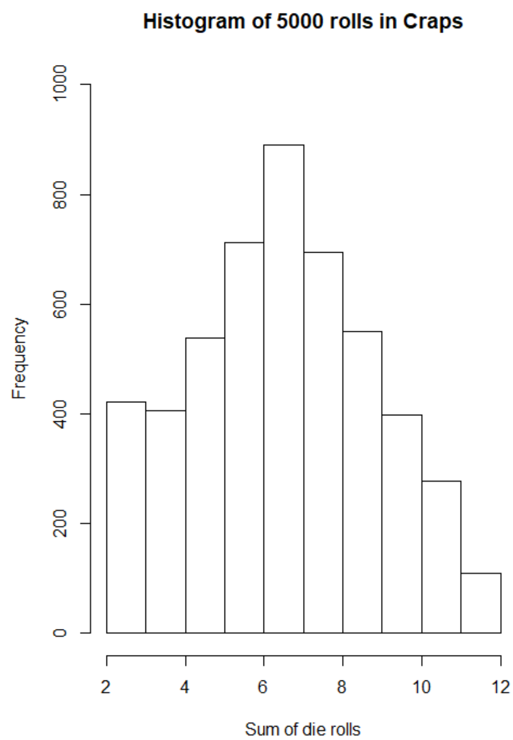
c.

I am concerned that the data is not linear and thus the data must be transformed to see a better fit. Perhaps the data fluctuates over time, so a linear plot will not show an accurate representation of the data.

3. Exercise 3

a. Chance that Adam doubles money in the first round: 2/9 => 22.2%. Chance that Adam loses all money in the first round: 1/9 => 11.1%

b. [1] 9 7 10 7 9



c.

d.
```
> occurrence_7  = sum(sum_outcome_dice == 7)
> occurrence_11 = sum(sum_outcome_dice == 11)
>  # Adam doubled his money
> Adam_doubled <- (occurrence_7 + occurrence_11)/length(sum_outcome_dice) * 100
>
> occurrence_2  = sum(sum_outcome_dice == 2)
> occurrence_3  = sum(sum_outcome_dice == 3)
> occurrence_12 = sum(sum_outcome_dice ==12)
>  # Adam lost all his money
> Adam_lost <- (occurrence_2 + occurrence_3 + occurrence_12)/length(sum_outcome_dice) * 10
0
> Adam_doubled
[1] 23.36
> Adam_lost
[1] 10.64
```

e. If event A is Adam winning and event B is Adam loosing, those events A and B would be mutually exclusive as Adam is unable to both win and lose on the same turn.

f. $P(A \cap B) \neq P(A)P(B)$: $Since\ P(A \cap B) = 0\ and\ P(A) = \frac{2}{9}\,; P(B) = \frac{1}{9}$

4. Exercise 4

a. n = 365; p = 0.40

b. mean = n*p = 365 * 0.40 = 146 in.; sd = sqrt(n*p*(1-p)) = sqrt(365*.4*.6) = 9.36 in.

c.
```
> dbinom(145, size = 356, prob = 0.4)
[1] 0.04133747
```

d.
```
> pbinom(175, size = 356, prob = 0.4) - pbinom(125, size = 356, prob = 0.4) + dbinom(175, size = 3
56, prob = 0.4)
[1] 0.9668713
```

e.
```
> 1 - pnorm(230, mean = 200, sd = 20)
[1] 0.0668072
```