Stats 10 Lab 3 Submission
Name: Enmin Zhou
UID: 104756697

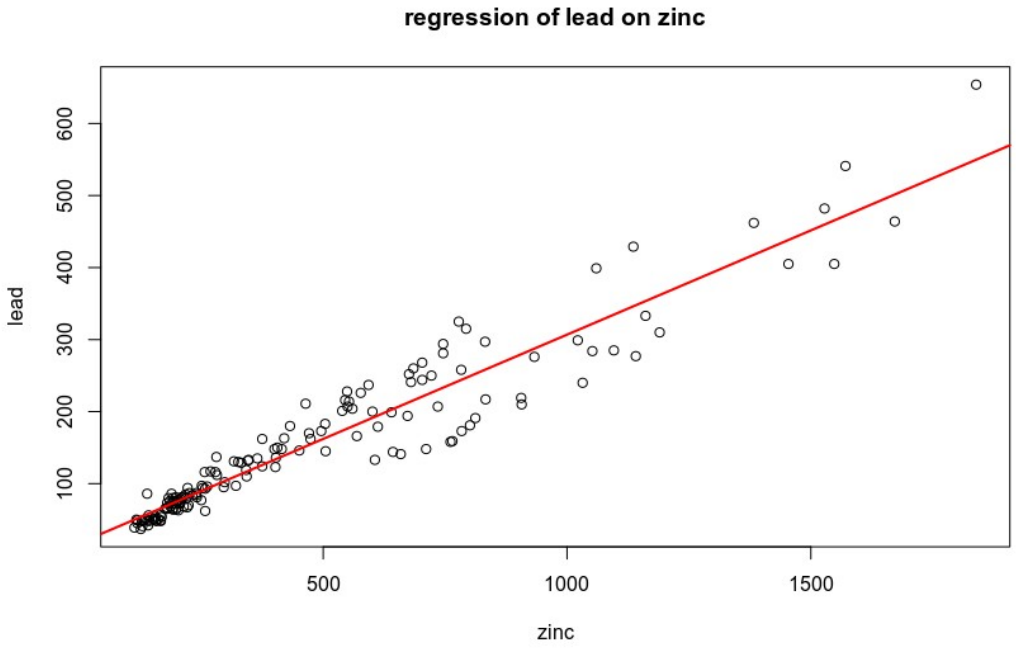
Exercise 1

1) a)

```
> soil<-read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/soil_complete.txt",
header=TRUE)
> linear_model <- lm(soil$lead ~ soil$zinc)
> summary(linear_model)

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
```

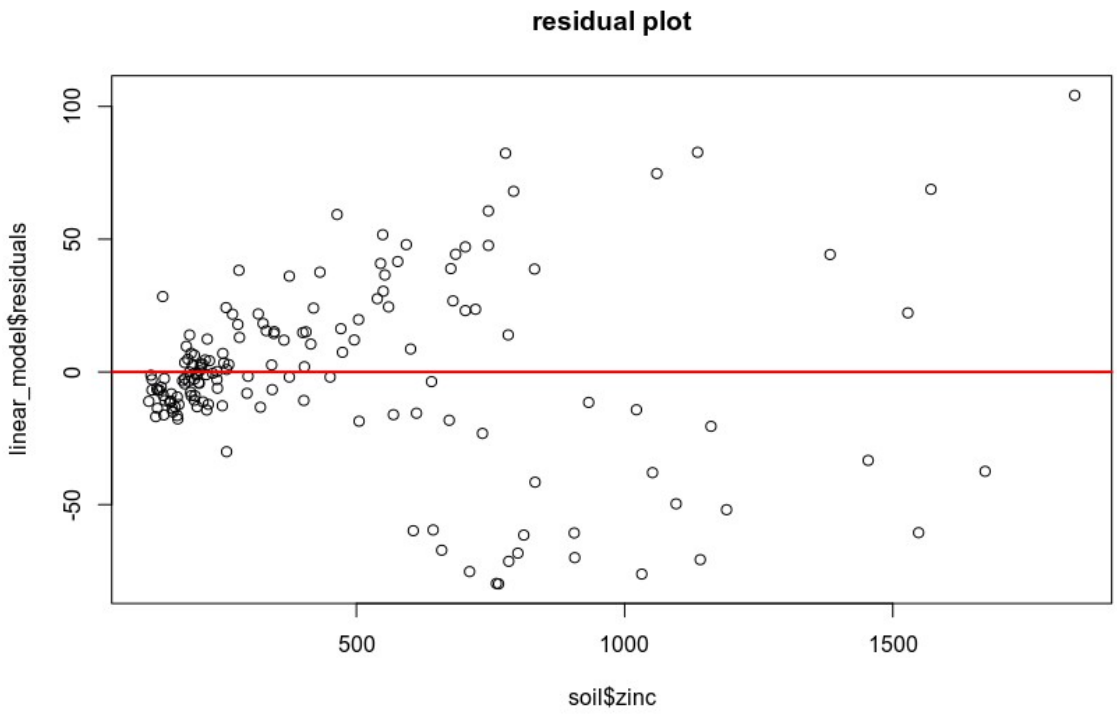
b)

```
> plot(soil$lead ~ soil$zinc, xlab="zinc", ylab="lead", main="regression of lead on zinc")
> abline(linear_model, col="red", lwd=2)
```



c)

```
> plot(linear_model$residuals ~ soil$zinc, main="residual plot")
> abline(a=0, b=0,col='red',lwd=2)
```



d)

lead=17.367688 + 0.289523*zinc

e)

17.36788+0.289523*1000 = 306.8909 ppm

f)

0.289523*100 = 28.9523 ppm higer

g)

0.91, it means that 91 percent of the change in the lead level can be explained by the zinc level.

h)

Linearity and symmetry assumptions are satisfied. The equal variance assumption is violated.

Exercise 2

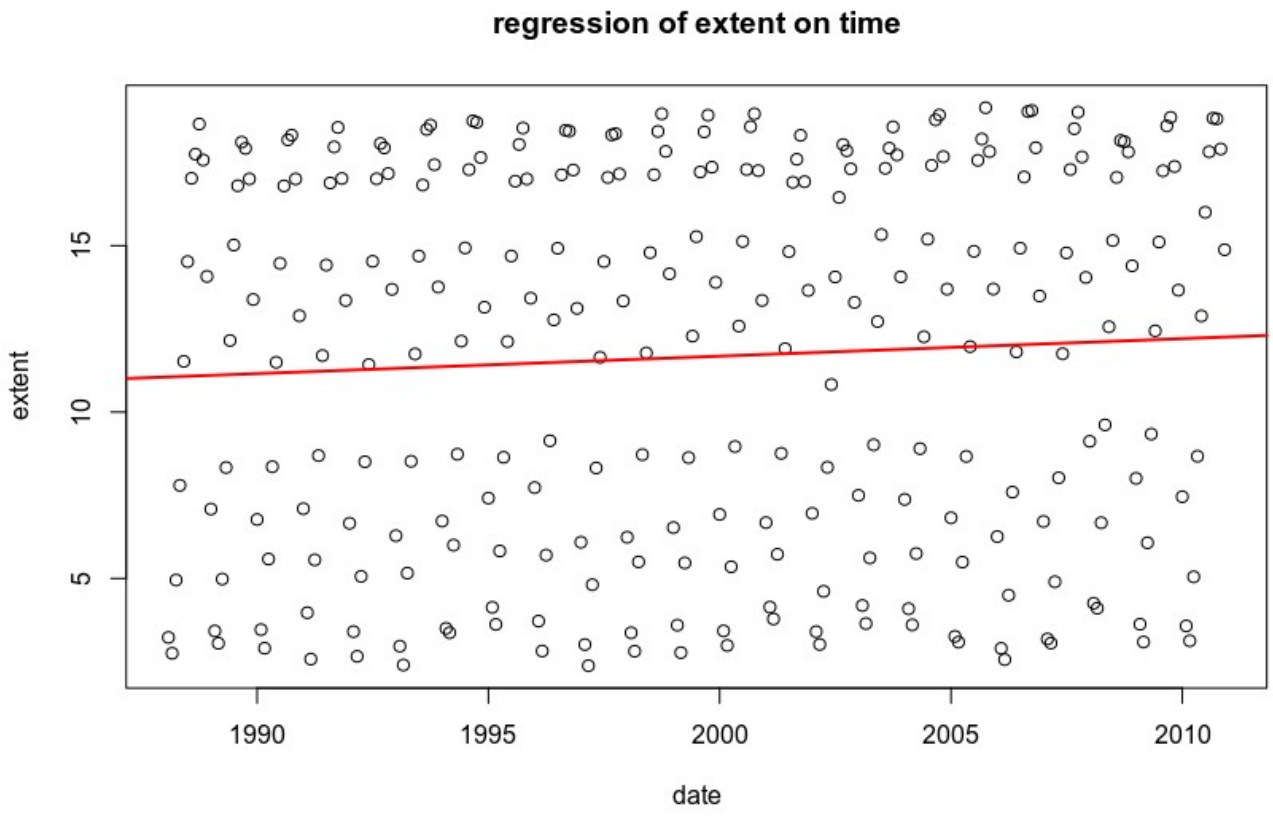
2) a)

```
> linear_model2 <- lm(ice$Extent ~ ice$Date)
> summary(linear_model2)

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)

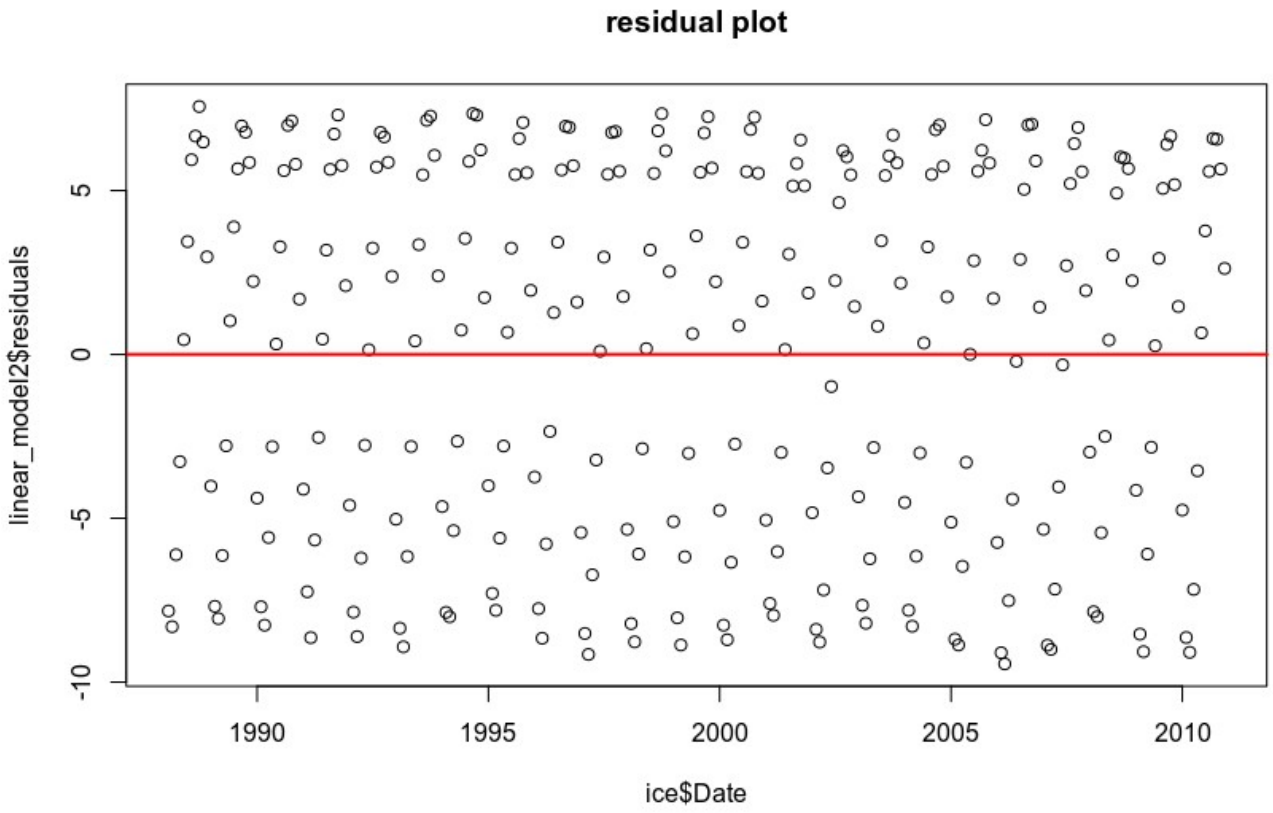
```
> plot(ice$Extent ~ ice$Date, xlab="date", ylab="extent", main="regression of extent on time")
> abline(linear_model2, col="red", lwd=2)
```



No.

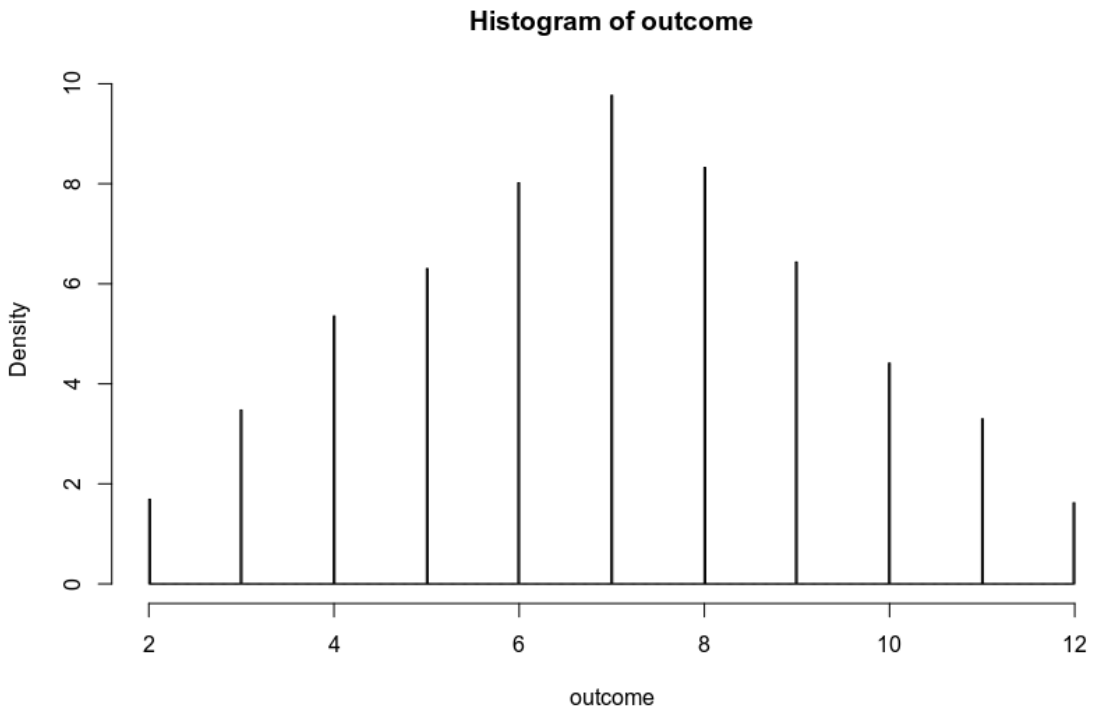
c)

```
> plot(linear_model2$residuals ~ ice$Date, main="residual plot")
> abline(a=0, b=0,col='red',lwd=2)
```



The linearity is violated. The symmetry and equal variance assumption are satisfied.
Exercise 3
3) a)
double: 1/6*1/6*3*2 + 1/6*1/6*2 = 2/9
lose: 1/6*1/6*2 + 1/6*1/6*2 + 1/6*1/6*2 = 1/6
b)

```
> set.seed(123)
> numbers = 1:6
> rand_dice = replicate(5000, sample(numbers, 2, replace = TRUE))
> outcome <- colSums(rand_dice)
> histogram(outcome)
```



c)

```
>> t<-table(outcome)
> double <- (t[6] + t[10])/5000
> lose <- (t[1]+t[2]+t[11])/5000
```

Double: 0.2226, Lose:0.1156

d)

These two events are disjoint since Adam cannot get (7,11) and (2,3,12) at the same time.

e)

P(double) = 2/9, P(lose) = 1/6, P(double & lose) = 0. Therefore they are not independent events.

Exercise 4

4)a)

p = 0.4, n = 365

b)

mean: np = 146, standard deviation: sqrt(np(1-p)) = 9.36

c)

```
> dbinom(145, size = 365, prob = 0.4)
[1] 0.04239996
```

Probability is 0.04239996

d)

```
> pbinom(175, size = 365, prob = 0.4) - pbinom(125, size = 365, prob = 0.4)
[1] 0.98543
```

Probability is 0.98543

e)

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

Probability is 0.0668702