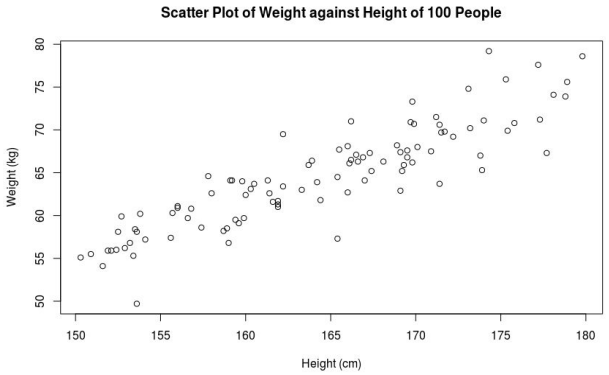
sap21 _ MAM5120 _ assessed practical 2 _ 6th December 2017

Question 1

a) `> plot(height, weight, main="Scatter Plot of Weight against Height of 100 People", ylab="Weight (kg)", xlab="Height (cm)")`



b) `> abline(lm(weight~height), col="red")`


```
> hw.lm <-lm(weight~height)
> summary(hw.lm)
Call:
lm(formula = weight ~ height)

Residuals:
    Min      1Q  Median      3Q     Max
-7.9750 -1.5062 -0.0422  1.6109  7.9519

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -45.73068    5.79994  -7.885 4.47e-12 ***
height        0.67113    0.03527  19.031  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.752 on 98 degrees of freedom
Multiple R-squared:  0.787,     Adjusted R-squared:  0.7849
F-statistic: 362.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
> a <- -45.73068
> b <- 0.67113
y = a + bx
y = -45.73068 + 0.67113(x)
```

c) what is meaning of Estimate in row labeled "height" in table of (b)

slope of the graph

Question 2

```
> ph <- c(5.3, 8.0, 6.0, 6.2, 6.7, 7.4, 7.7, 7.5, 6.6, 5.1, 7.0, 7.7)
> eff <- c(30.2, 59, 37.3, 38, 42, 52, 56.3, 55.4, 43.3, 27.8, 44.8, 55.1)
```

a)

```
> pe.lm1 <- lm(eff~ph)
> summary(pe.lm1)
Call:
lm(formula = eff ~ ph)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8341 -0.3387  0.2636  0.8577  2.3357

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.3889     3.2832  -8.647 5.92e-06 ***
ph           10.8604     0.4808  22.588 6.51e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.53 on 10 degrees of freedom
Multiple R-squared:  0.9808,    Adjusted R-squared:  0.9789
F-statistic: 510.2 on 1 and 10 DF,  p-value: 6.51e-10
```

Model1 <- “y = -28.3889 + 10.8604(x)”

```
> pe.lm2 <- lm(eff~ph+(I(ph^2)))
> summary(pe.lm2)
Call:
lm(formula = eff ~ ph + (I(ph^2)))

Residuals:
    Min      1Q  Median      3Q     Max
-2.10391 -0.62425  0.03311  0.70171  2.33325

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.0767    21.7334   0.740   0.4783
ph           -2.9898     6.7240  -0.445   0.6671
I(ph^2)       1.0562     0.5118   2.064   0.0691 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.329 on 9 degrees of freedom
Multiple R-squared:  0.987,     Adjusted R-squared:  0.9841
F-statistic: 340.4 on 2 and 9 DF,  p-value: 3.31e-09
```

Model2 <- “y <- 16.0767 - 2.9898(x) + 1.0562(x)”

Model 2 better fit as R-squared is higher

b)

> m1.r2 <- 0.9808
> m2.r2 <- 0.987

c)

> h0 <- "no significant difference"
> ha <- "some significant difference"

```
> anova(pe.lm1, pe.lm2)
Analysis of Variance Table

Model 1: eff ~ ph
Model 2: eff ~ ph + (I(ph^2))
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     10 23.409
2      9 15.890  1    7.5198 4.2592 0.06905 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

> p-value <- 0.06905

no need to reject null hypothesis as p-value (at 10%) is higher than 5%

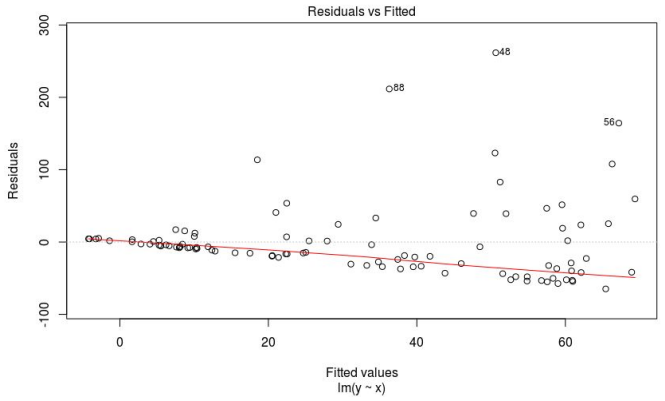
Question 3

```
> A.df <- read.csv("Question3_Week9.csv")
```

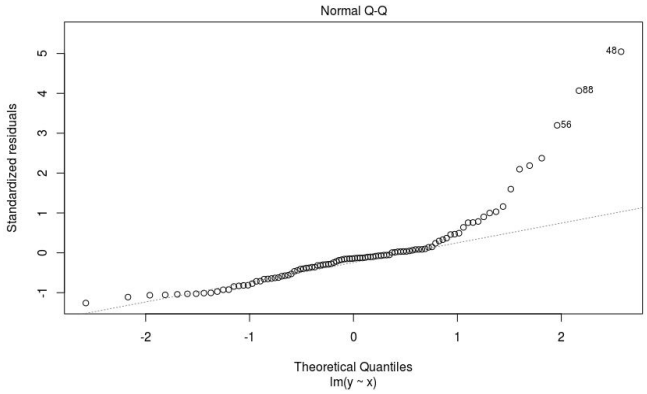
a)

```
> a.lm <- lm(y~x, data = A.df)
> summary(a.lm)
Call:
lm(formula = y ~ x, data = A.df)

Residuals:
    Min      1Q  Median      3Q     Max
-64.828 -30.008  -7.297   4.606 261.626

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -4.314      9.896  -0.436    0.664
x             14.830      3.426   4.328 3.63e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52.3 on 98 degrees of freedom
Multiple R-squared:  0.1605,    Adjusted R-squared:  0.1519
F-statistic: 18.73 on 1 and 98 DF,  p-value: 3.628e-05
```

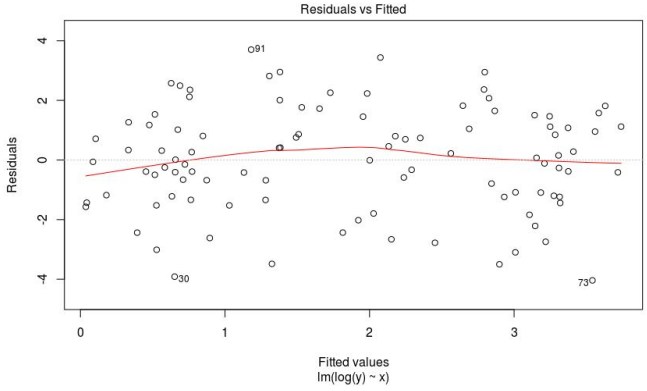
```
> plot(a.lm)
```

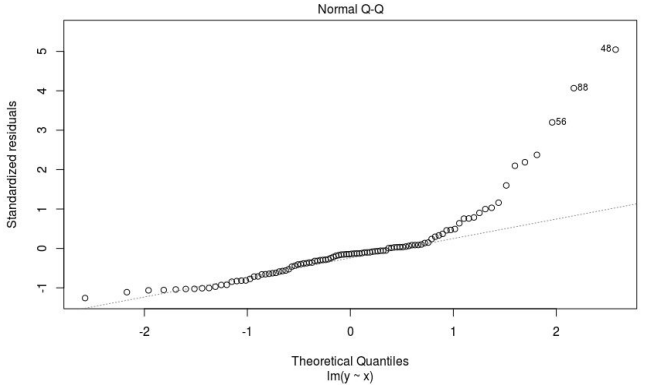


R code: https://github.com/sap218/R/blob/master/mam5120/assessed_prac_2/assessed_prac_2.R



b) `> alog.lm <- lm(log(y)~x, data = A.df)`
`plot(alog.lm)`



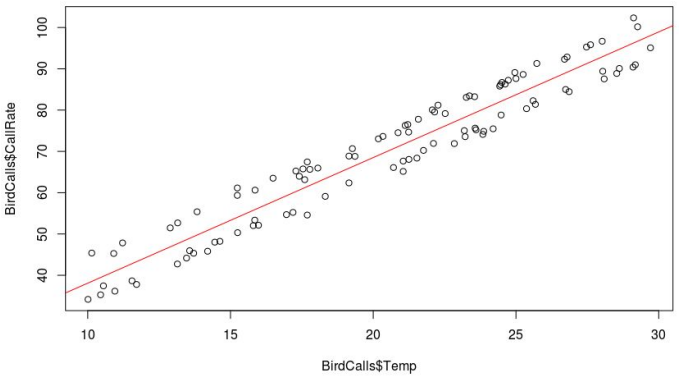


c) `log(y) = a + bx`
d) `each time x is increased by 1, y is multiplied by` e^{bx}

Question 4

c)

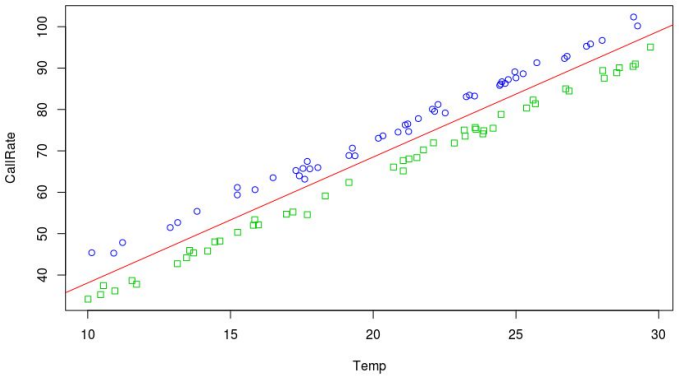
```
> BirdCalls <-
read.csv("~/git/R/R/mam5120/assessed_prac_2/BirdCalls.csv")
> plot(BirdCalls$Temp, BirdCalls$CallRate)
> abline(lm(BirdCalls$CallRate~BirdCalls$Temp), col="red")
```




```
> bc.lm <- lm(CallRate~Temp, data = BirdCalls)
> summary(bc.lm)
Call:
lm(formula = CallRate ~ Temp, data = BirdCalls)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8980 -4.2658 -0.1974  4.3263  7.1140

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.68117    1.78674   4.299 4.05e-05 ***
Temp         3.04165    0.08432  36.073  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.56 on 98 degrees of freedom
Multiple R-squared:   0.93,     Adjusted R-squared:  0.9292
F-statistic:  1301 on 1 and 98 DF,  p-value: < 2.2e-16
```

d)

```
> plot(CallRate~Temp, data = BirdCalls, pch=Species, col=Species+3)
> abline(lm(BirdCalls$CallRate~BirdCalls$Temp), col="red")
```

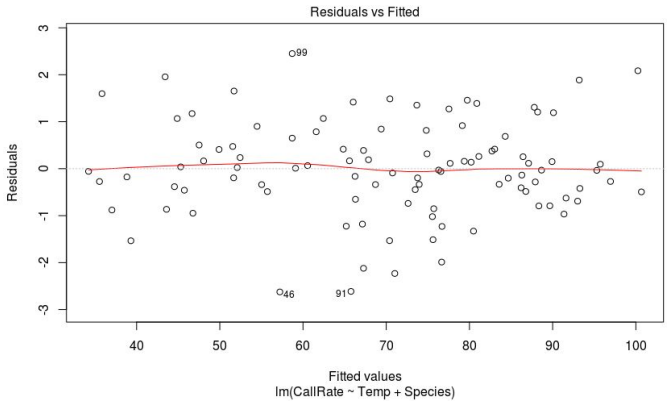


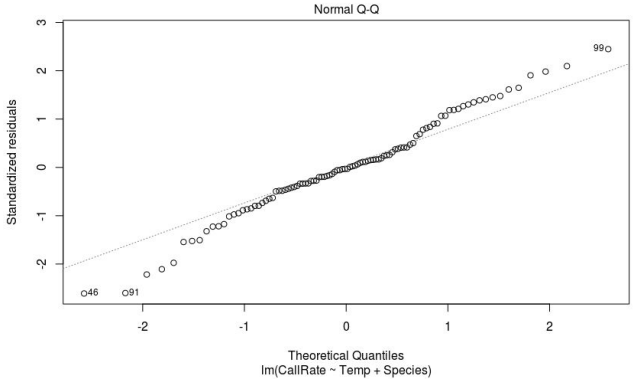
e) `> bc.lm2 <- lm(CallRate~Temp+Species, data = BirdCalls)`
`> summary(bc.lm2)`

```
Call:
lm(formula = CallRate ~ Temp + Species, data = BirdCalls)

Residuals:
     Min       1Q   Median       3Q      Max
-2.62449 -0.48942 -0.03041  0.54017  2.45042

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.26512    0.40630   10.50   <2e-16 ***
Temp         2.99318    0.01884  158.86   <2e-16 ***
Species      8.81838    0.20378   43.27   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.017 on 97 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9965
F-statistic: 1.401e+04 on 2 and 97 DF,  p-value: < 2.2e-16
```

`> plot(bc.lm2)`

R code: https://github.com/sap218/R/blob/master/mam5120/assessed_prac_2/assessed_prac_2.R





species estimate is quite high

f)
```
> BirdCalls3.lm <- lm(CallRate~Temp*Species, data = BirdCalls)
> summary(BirdCalls3.lm)

Call:
lm(formula = CallRate ~ Temp * Species, data = BirdCalls)

Residuals:
     Min       1Q   Median       3Q      Max
-2.67998 -0.48657  0.03269  0.54759  2.33505

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.97674    0.51499   7.722  1.1e-11 ***
Temp          3.00748    0.02452 122.664  < 2e-16 ***
Species       9.53778    0.81427  11.713  < 2e-16 ***
Temp:Species -0.03501    0.03836  -0.913    0.364
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.018 on 96 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9965
F-statistic:  9326 on 3 and 96 DF,  p-value: < 2.2e-16
```

p-value for Temp:Species is 0.364: R-squared is 0.9966 with adjusted R-squared being 0.9965

g)
```
> rs.m2 <- c(-2.62449, -0.48942, -0.03041,  0.54017,  2.45042)
> rs.m3 <- c(-2.67998, -0.48657,  0.03269,  0.54759,  2.33505)
> rs.m2.sum <- sum(rs.m2)
> rs.m3.sum <- sum(rs.m3)
> rs.m2.sum
[1] -0.15373
> rs.m3.sum
[1] -0.25122
```

model 2 produced -0.15373 sum of residual squares, model 3 produced -0.25122 sum of residual squares

R code: https://github.com/sap218/R/blob/master/mam5120/assessed_prac_2/assessed_prac_2.R