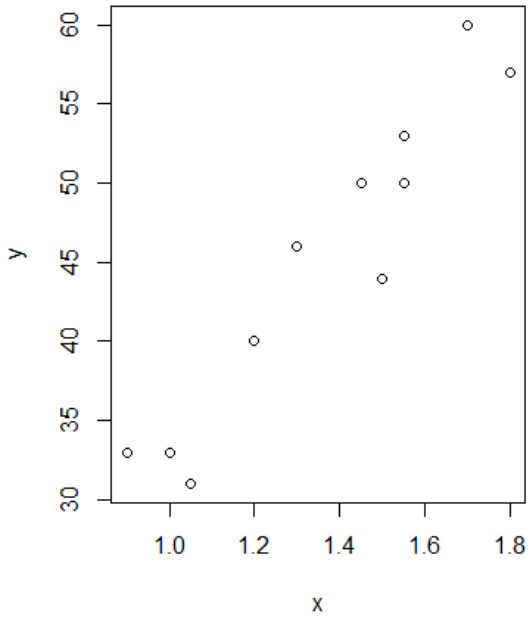
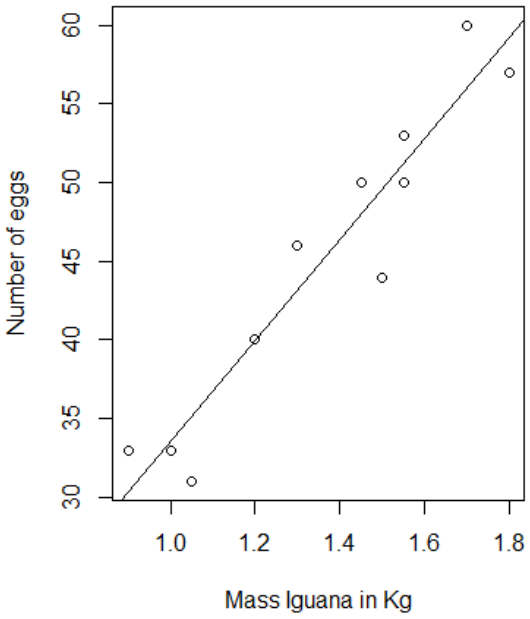
STA2023 R lab 8 Correlation and Regression

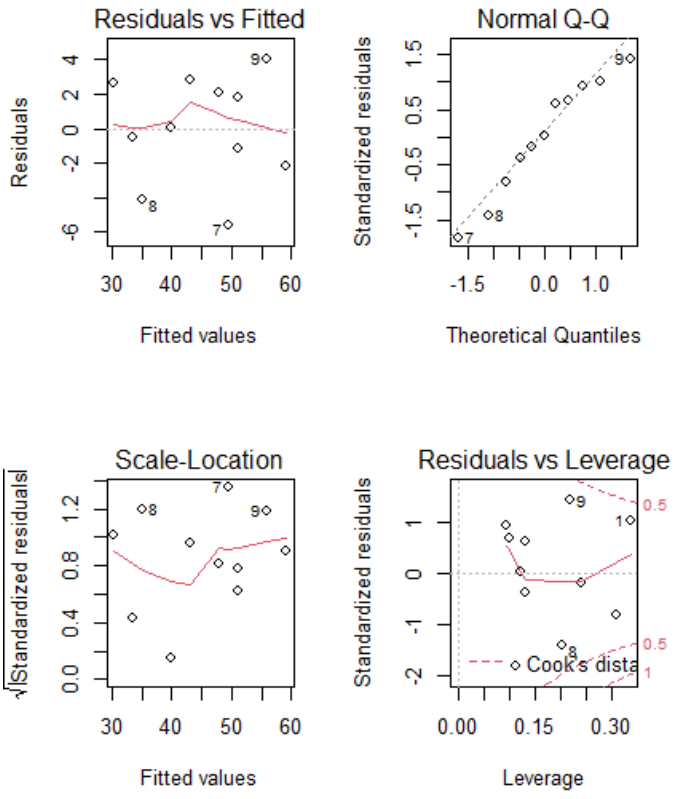
```
> #correlation
> # weight of female iguanas and number of eggs produced (source; Intro to Bi
ological Statistics, by Hampton, R.)
> #y, number of eggs; x, mass of iguana in Kg
> x <- c(0.90,1.55,1.30,1.00, 1.55, 1.80,1.50, 1.05,1.70,1.20, 1.45) # 11 dat
a values
> y <- c(33,50,46,33, 53,57,44,31,60,40,50)
> cor(x,y)
[1] 0.9518183
> plot(x,y) # just the plot, no regression line
```



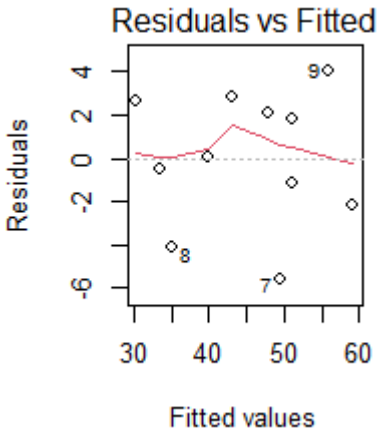

```
> plot(x,y, xlab="Mass Iguana in Kg", ylab = "Number of eggs")
> model1 <- lm(y~x)
> abline(model1) # it draws the regression line
```



```
> summary(model1)
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-5.5545 -1.6661  0.0654  2.3671  4.0322

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.455      4.796   0.303    0.769
x             32.066      3.444   9.311 6.46e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.222 on 9 degrees of freedom
Multiple R-squared:  0.906,    Adjusted R-squared:  0.8955
F-statistic:  86.7 on 1 and 9 DF,  p-value: 6.459e-06

> coef(model1)
(Intercept)           x
   1.454829   32.066459

> # equation: y(number of eggs)=1.455 + 32.066 x (weight)

> # p-value: 6.459e-06
> # recall: we test the null hypothesis of no linear correlation in populatio
n against the alternative that there is linear correlation present.

> # diagnostic plots:
> par(mfrow = c(2, 2))
> plot(model1)
```



```
> plot(model1, 1) # only the Residual vs fitted produced.
On this graph, ideally, points should be distributed symmetrically around y=0
```




```
> # predicting values for given input of x:
> y=1.454829 + 32.066459 * 1.1    # using the regression equation
> y
[1] 36.72793

> predict(model1, newdata=data.frame(x=1.1)) # prediction using r function
       1
36.72793

> predict(model1, newdata=data.frame(x=c(0,1.1))) # predictions for two x
values: 1 corresponds to zero, and 2 correspond to 1.1
        1         2
 1.454829 36.727934

> # Notice that if no significant corr is found, the model cannot be used to
make predictions.
```