Report Sample

Student name

3/2/2021

Introduction

Kernel regression is a non-parametric estimator that estimates the conditional expectation of two variables which is random. The goal of a kernel regression is to discover the non-linear relationship between two random variables. To discover the non-linear relationship, kernel estimator or kernel smoothing is the main method to estimate the curve for non-parametric statistics. In kernel estimator, weight function is known as kernel function (R Core Team 2019). # Methods

Application using R

A study was conducted to determine how to maximize Mermaid meadowfoam flower production for extraction as vegetable oil.

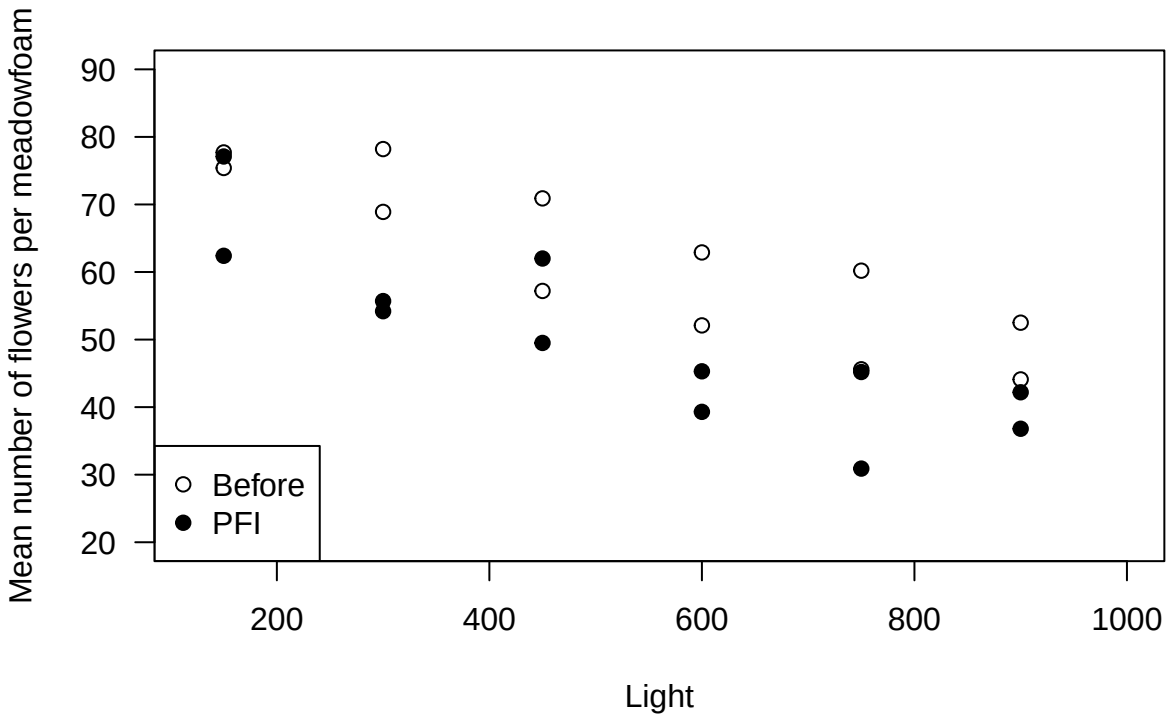
```
# loading package
library(GLMsData)
library(DT)
data(flowers)
# to display the data set from the loaded package
datatable(flowers)
```

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, pleas
```

```
There are 3 variables and 24 observations in flowers data set. The response variable is the mean number
```

1.Plot the average number of flowers produced per plant against the light intensity, distinguishing the two timings. Comment.

```
# Plotting
plot(Flowers ~ Light,data=flowers, pch=ifelse(Timing=="Before",1,19),las=1,
     xlab="Light", ylab="Mean number of flowers per meadowfoam",
     xlim=c(120, 1000), ylim=c(20, 90))
legend("bottomleft", pch=c(1,19), legend=c("Before","PFI"))
```




```
Comment here...
```

2.Suppose a model with the systematic component Flowers ~ Light + Timing was needed to model the data. What would such a systematic component imply about the relationship between the variables?

```
Answer here...
```

3.Suppose a model with the systematic component Flowers ~ Light + Timing + Light: Timing was needed to model the data. What would such a systematic component imply about the relationship between the variables?

```
Answer here...
```

4.Fit the two linear regression models with the systematic components specified above. Give the regression tables and comment.

```
# Fit model without interaction
m1=lm(Flowers~Light + Timing,data = flowers)
summary(m1)
```

```
##
## Call:
## lm(formula = Flowers ~ Light + Timing, data = flowers)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.040  -3.930  -1.819   5.587  11.866
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.399167   3.289525  25.353  < 2e-16 ***
## Light        -0.040490   0.005157  -7.851 1.11e-07 ***
## TimingPFI   -12.091667   2.642210  -4.576 0.000164 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.472 on 21 degrees of freedom
## Multiple R-squared:  0.7973, Adjusted R-squared:  0.778
## F-statistic: 41.29 on 2 and 21 DF,  p-value: 5.279e-08
```

comment here...

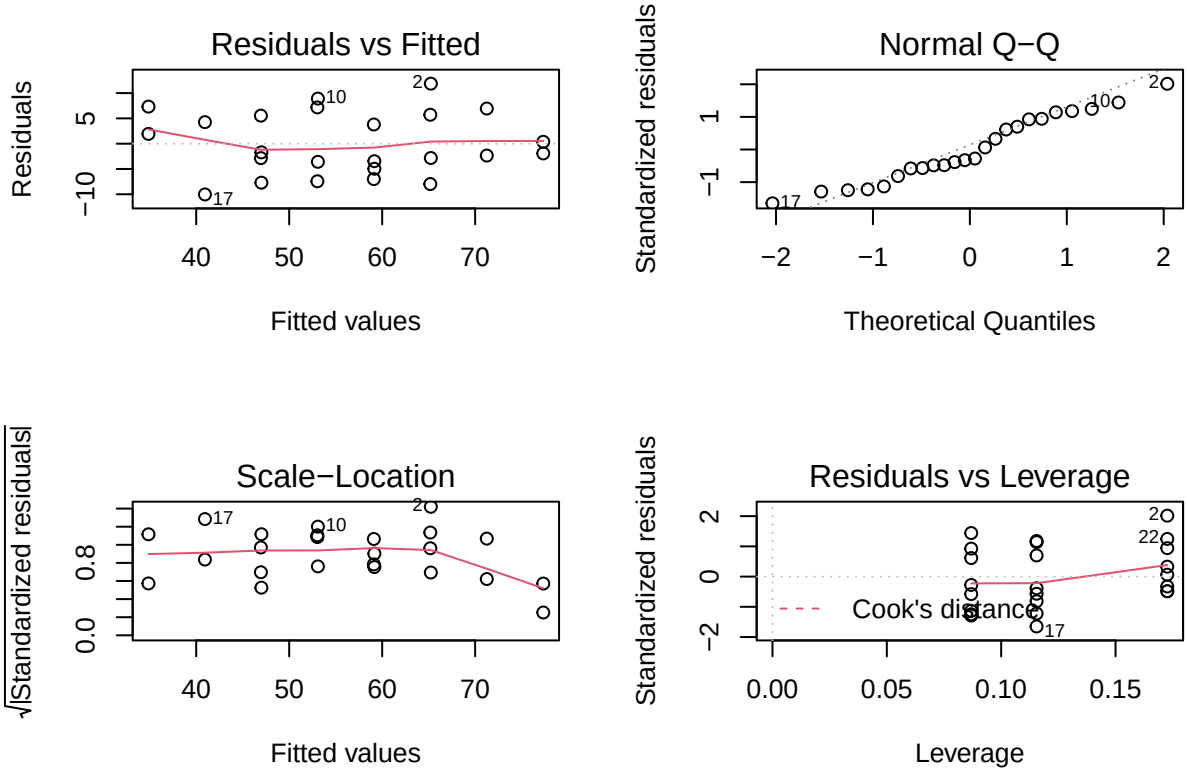
```
# Fit model with interaction
m2=lm(Flowers~Light*Timing,data = flowers)
summary(m2)
```

```
##
## Call:
## lm(formula = Flowers ~ Light * Timing, data = flowers)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.919 -3.991 -1.698  5.446 11.664
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      83.116667   4.364515  19.044 2.75e-14 ***
## Light            -0.039952   0.007471  -5.347 3.11e-05 ***
## TimingPFI       -11.526667   6.172356  -1.867   0.0766 .
## Light:TimingPFI  -0.001076   0.010566  -0.102   0.9199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.63 on 20 degrees of freedom
## Multiple R-squared:  0.7974, Adjusted R-squared:  0.767
## F-statistic: 26.24 on 3 and 20 DF,  p-value: 3.898e-07
```

Comment here...

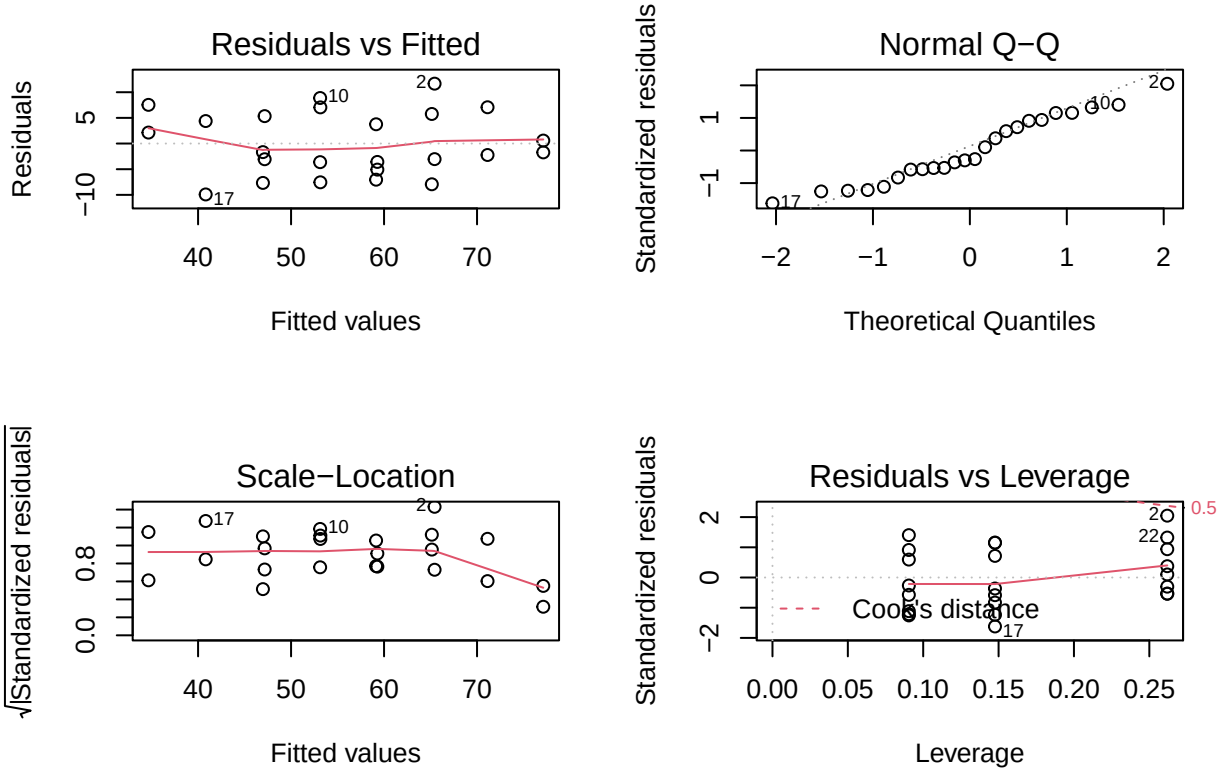
5.Perform a diagnostic analysis on the fitted linear regression models. Present the residuals plots and comment.

```
# Model without interaction
par(mfrow=c(2,2))
plot(m1)
```




```
  comment here ...
```

```
# Model with interaction
par(mfrow = c(2,2))
plot(m2)
```



Comment here ...

6.Identify any influential observations or outliers?

```
# Model without interaction
rowSums(influence.measures(m1)$is.inf)
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

Comment here ...

```
# Model with interaction
rowSums(influence.measures(m2)$is.inf)
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
##  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0
```

Comment here ...

7.What is the best model? Why?

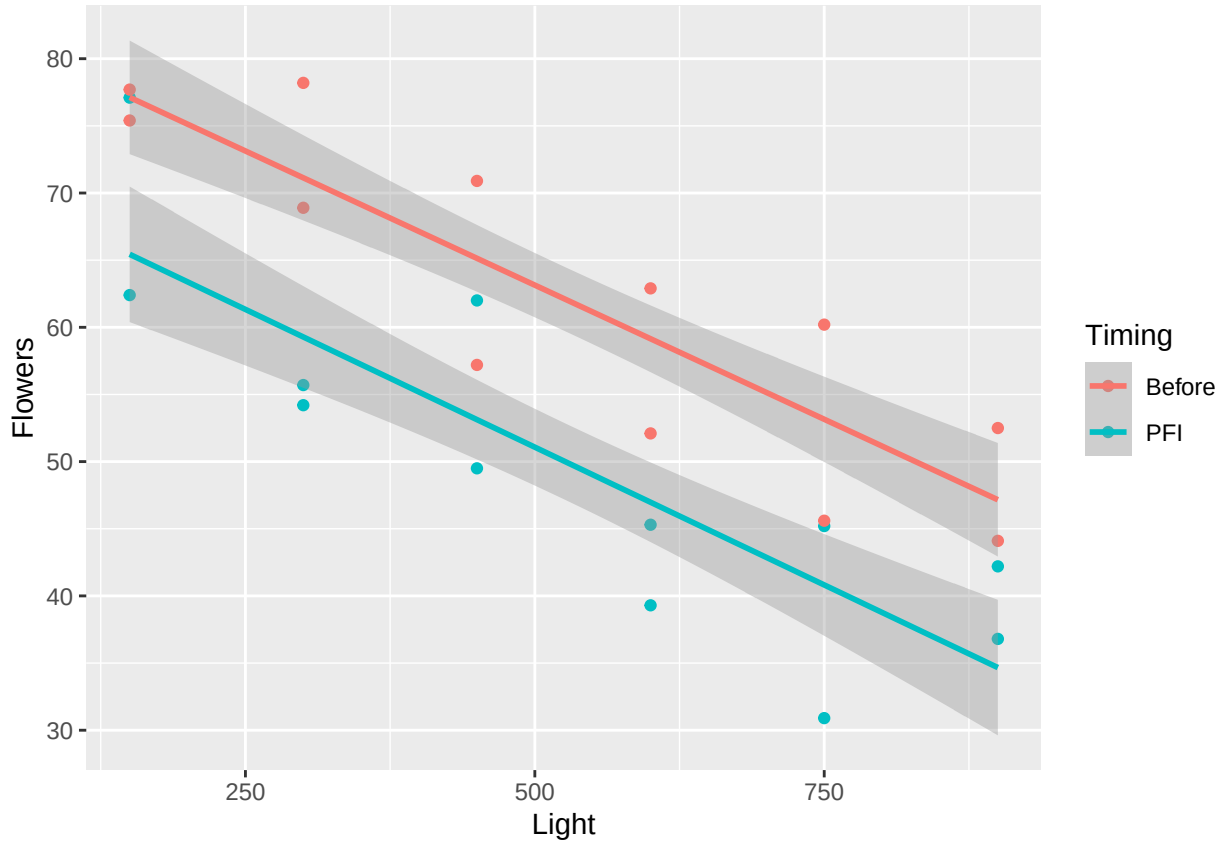
```r
# comparing models
anova(m1,m2)
```

```
## Analysis of Variance Table
##
## Model 1: Flowers ~ Light + Timing
## Model 2: Flowers ~ Light * Timing
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     21 879.64
## 2     20 879.18  1   0.45604 0.0104 0.9199
```

comment here...

8.Interpret the final model (best model)

```r
library(ggplot2)
# plot regression lines
ggplot(flowers,aes(x=Light,y=Flowers,col=Timing))+geom_point()+geom_smooth(method = "lm", formula = y~x
```



The results indicate that the light intensity is significant
after adjusting for the other variables.
The coefficient is negative suggesting that greater light intensity
is associated with lower mean number of flowers
If there were interaction this interpretation is only valid for Timing == Before

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.