Assignment 6 STAT 315-463: Multivariable Statistical Methods and Applications

Lisa Lu 31088272

2023-05-27

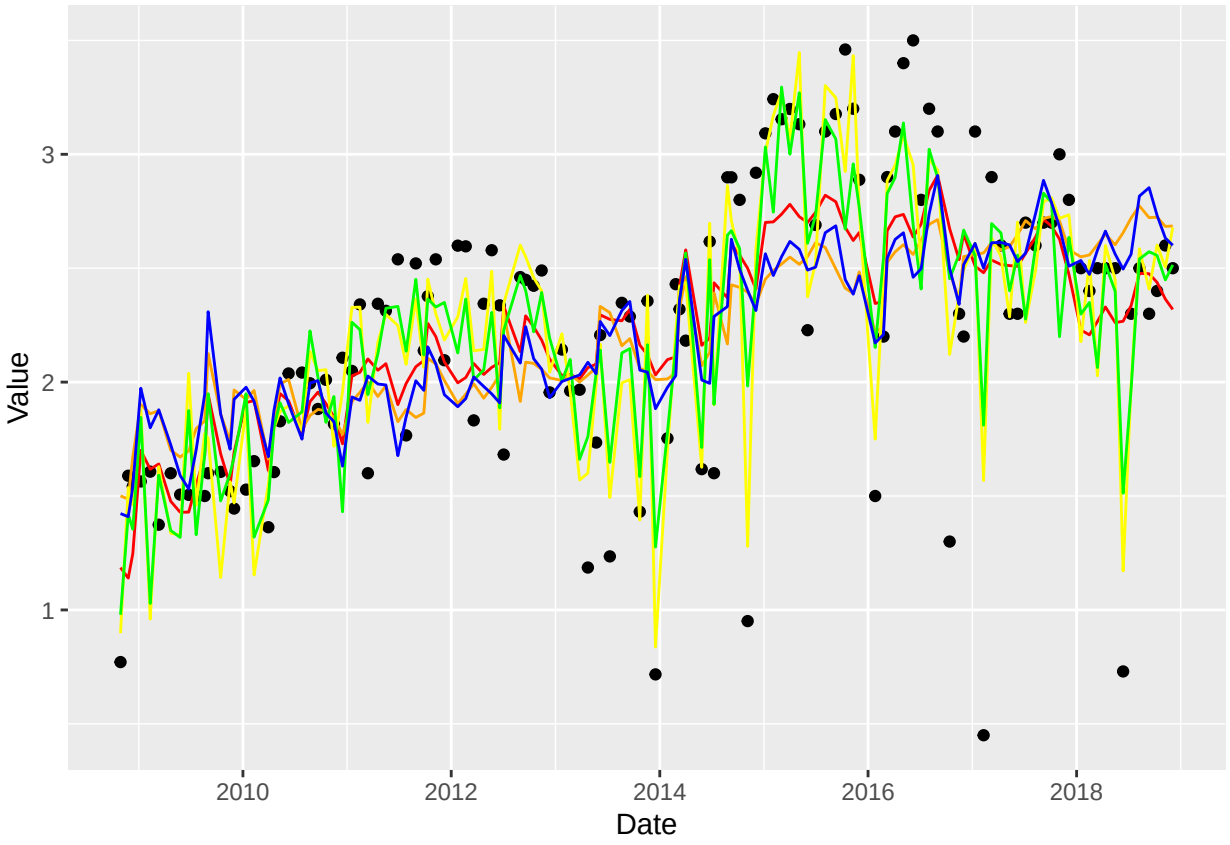
QUESTION 1 Generalised additive models

a) Import the data to R and fit a series of GAMs to the Value using a smoother on Date.

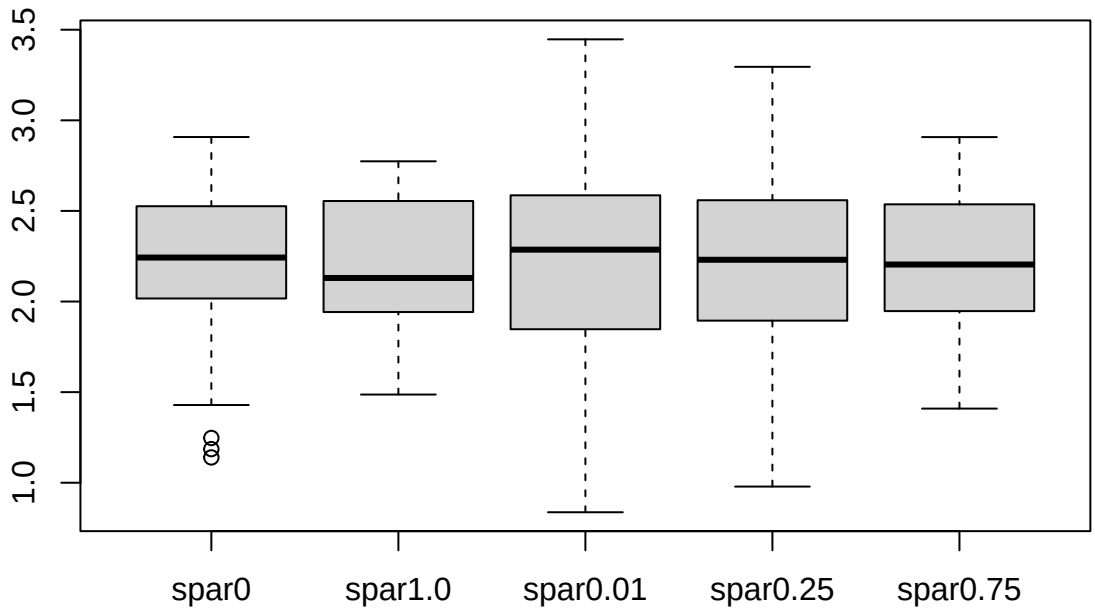
```
# Read in the datasets and convert the string "Date" into Date datatype variables
CCC05 <- read.table("CCC05.csv", header = TRUE, sep = ',', na.strings = "na")
CCC05$Date <- as.Date(CCC05$Date, "%d/%m/%Y")
```

```
library(gam)
library(ggplot2)

# Kept showing Error in names(dat) <- object$term :
# 'names' attribute [1] must be the same length as the vector [0]
CCC05 <- transform(CCC05, ndate = as.numeric(Date),
                   nyear  = as.numeric(format(Date, '%Y')),
                   nmonth = as.numeric(format(Date, '%m')),
                   nday = as.numeric(format(Date, '%j')))

# Start with the default model and 4 more with different spar parameters

CCC05.gam <- gam(Value ~ s(nyear) + s(nmonth) +s(nday), data = CCC05 )
pred_default <- predict(CCC05.gam)

CCC05.gam1 <- gam(Value ~ s(nyear, sp=1.0) + s(nmonth, sp = 1.0) +
                    s(nday, sp = 1.0), data = CCC05)
pred1 <- predict(CCC05.gam1)

CCC05.gam2 <- gam(Value ~ s(nyear, sp=0.01) + s(nmonth, sp = 0.01) +
                    s(nday, sp = 0.01), data = CCC05)
pred2 <- predict(CCC05.gam2)

CCC05.gam3 <- gam(Value ~ s(nyear, sp=0.25) + s(nmonth, sp = 0.25) +
                    s(nday, sp = 0.25), data = CCC05)
pred3 <- predict(CCC05.gam3)

CCC05.gam4 <- gam(Value ~ s(nyear, sp=0.75) + s(nmonth, sp = 0.75) +
                    s(nday, sp = 0.75), data = CCC05)
pred4 <- predict(CCC05.gam4)
```

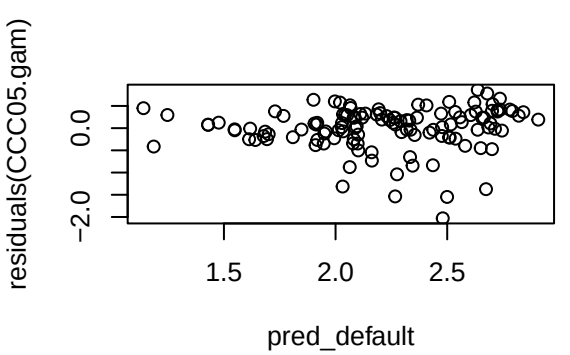
```
ggplot(data = CCC05, aes(x = Date, y = Value)) +
  geom_point() +
  geom_line(aes(x=Date, y=pred_default), colour = "red") +
  geom_line(aes(x=Date, y=pred1), colour = "orange") +
  geom_line(aes(x=Date, y=pred2), colour = "yellow") +
  geom_line(aes(x=Date, y=pred3), colour = "green") +
  geom_line(aes(x=Date, y=pred4), colour = "blue")
```

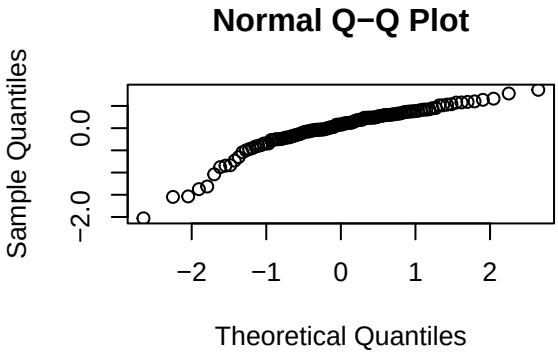


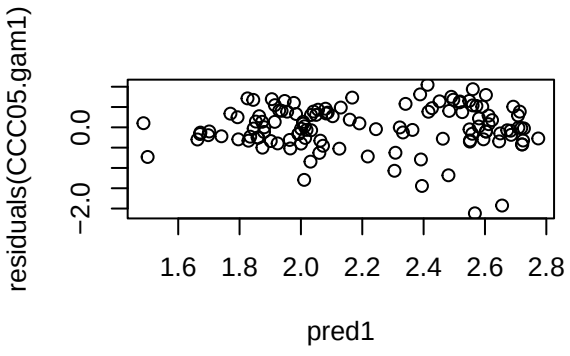

```
data <- data.frame(spar0 = fitted.values(CCC05.gam),
                   spar1.0 = fitted.values(CCC05.gam1),
                   spar0.01 = fitted.values(CCC05.gam2),
                   spar0.25 = fitted.values(CCC05.gam3),
                   spar0.75 = fitted.values(CCC05.gam4)
                   )

boxplot(data)
```

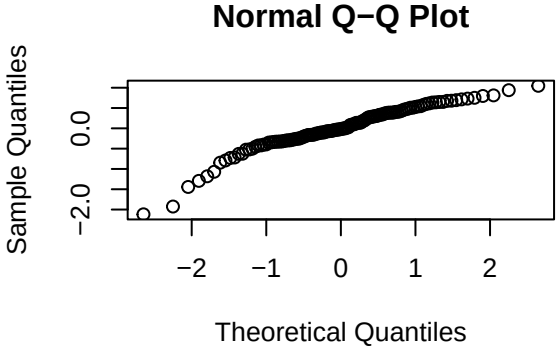



```
par(mfrow=c(2,2))
plot(pred_default, residuals(CCC05.gam),)
qqnorm(residuals(CCC05.gam))
plot(pred1, residuals(CCC05.gam1))
qqnorm(residuals(CCC05.gam1))
```

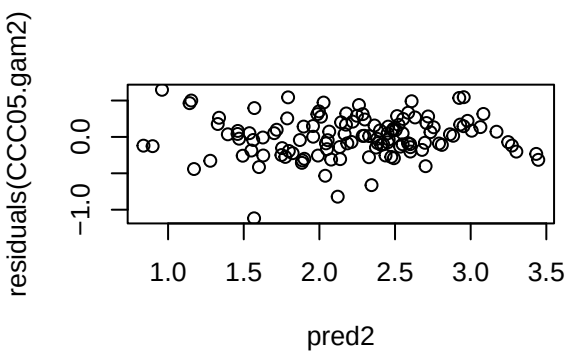


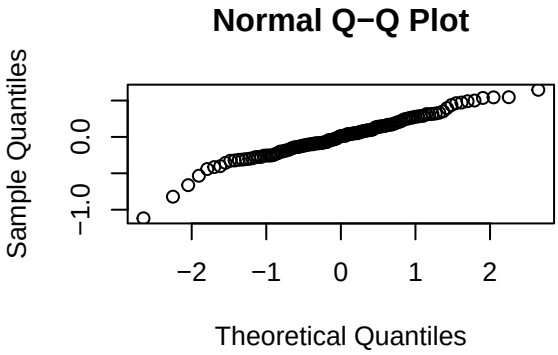


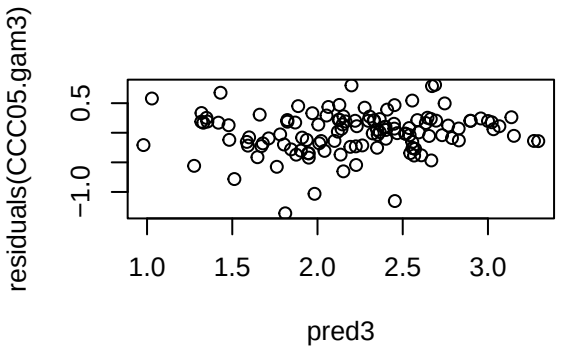


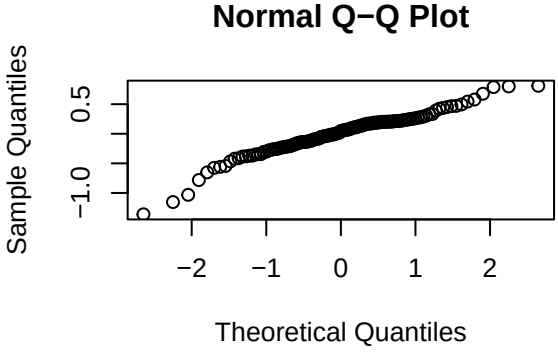



```
plot(pred2, residuals(CCC05.gam2))
qqnorm(residuals(CCC05.gam2))
plot(pred3, residuals(CCC05.gam3))
qqnorm(residuals(CCC05.gam3))
```

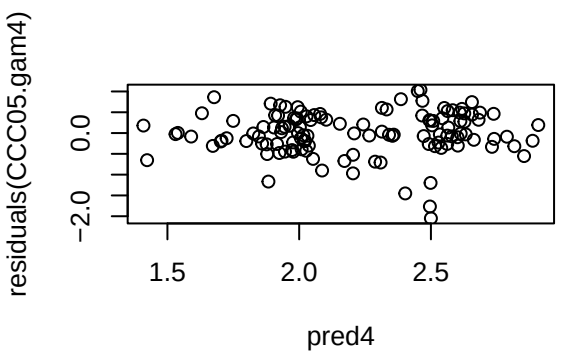


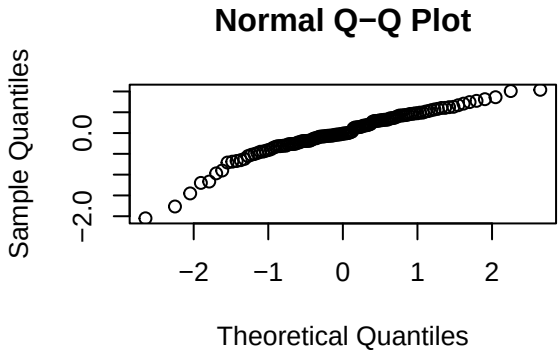







```
plot(pred4, residuals(CCC05.gam4))
qqnorm(residuals(CCC05.gam4))
```






```
ECAN93 <- read.table("ECAN93.csv", header = TRUE, sep = ',', na.strings = "na")
ECAN93$Date <- as.Date(ECAN93$Date, "%d/%m/%Y")

ECAN93 <- transform(ECAN93, ndate = as.numeric(Date),
                    nyear  = as.numeric(format(Date, '%Y')),
                    nmonth = as.numeric(format(Date, '%m')),
                    nday = as.numeric(format(Date, '%j')))


# Start with the default model and 4 more with different spar parameters
```

QUESTION 2 Multiple Comparisons

(a) Carry out an analysis of variance on the data with Herbicide as the explanatory variable and Grass_percent" as the response.
(b) Discuss the residuals
(c) Carry out an LSD type analysis comparing all possible pairs of treatments. Note which pairs have a significant difference.
(d) Carry out pairwise comparisons using Bonferroni, Tukey and Dunnett adjustments and in each case show the pairs with significant differences.
(e) How do the conclusions in (c) and (d) differ?