

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

Lisa Lu 31088272

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)
summary(CCC05.gam)
```

```
##
## Call: gam(formula = Value ~ s(nyear) + s(nmonth) + s(nday), data = CCC05)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.03005 -0.19084  0.08084  0.30188  0.86313
##
## (Dispersion Parameter for gaussian family taken to be 0.2553)
##
##      Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 28.0784 on 110.0003 degrees of freedom
## AIC: 195.3644
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## s(nyear)      1 12.0657 12.0657 47.2689 3.926e-10 ***
## s(nmonth)     1  0.2207  0.2207  0.8646   0.3545
## s(nday)       1  1.3653  1.3653  5.3487   0.0226 *
## Residuals 110 28.0784  0.2553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##           Npar Df Npar F      Pr(F)
## (Intercept)
## s(nyear)           3 7.8969 8.091e-05 ***
## s(nmonth)          3 1.7100   0.1691
## s(nday)            3 1.8543   0.1416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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)
summary(CCC05.gam1)
```

```
##
## Call: gam(formula = Value ~ s(nyear, sp = 1) + s(nmonth, sp = 1) +
##           s(nday, sp = 1), data = CCC05)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.11710 -0.28785 -0.01818  0.38619  1.04720
##
## (Dispersion Parameter for gaussian family taken to be 0.2981)
##
## Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 35.0015 on 117.3956 degrees of freedom
## AIC: 207.6818
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##           Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 1)  1.0 11.932  11.9322 40.0210 4.736e-09 ***
## s(nmonth, sp = 1) 1.0  0.111   0.1113  0.3732  0.54243
## s(nday, sp = 1)   1.0  1.666   1.6661  5.5882  0.01972 *
## Residuals        117.4 35.001   0.2981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##           Npar Df  Npar F      Pr(F)
## (Intercept)
## s(nyear, sp = 1)      0.0 17.4747 0.003987 **
## s(nmonth, sp = 1)     0.0  0.2069 0.015504 *
## s(nday, sp = 1)       1.6  1.7826 0.179368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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)
summary(CCC05.gam2)
```

```
##
## Call: gam(formula = Value ~ s(nyear, sp = 0.01) + s(nmonth, sp = 0.01) +
##       s(nday, sp = 0.01), data = CCC05)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1183 -0.1609  0.0117  0.1688  0.6473
##
## (Dispersion Parameter for gaussian family taken to be 0.2468)
##
## Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 9.5027 on 38.4959 degrees of freedom
## AIC: 205.1122
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 0.01)    1.000  9.5682   9.5682 38.7614 2.649e-07 ***
## s(nmonth, sp = 0.01)    1.000  0.1811   0.1811  0.7337  0.39699
## s(nday, sp = 0.01)      1.000  0.7149   0.7149  2.8960  0.09686 .
## Residuals              38.496  9.5027   0.2468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##              Npar Df Npar F      Pr(F)
## (Intercept)
## s(nyear, sp = 0.01)      8.9  3.585  0.002615 **
## s(nmonth, sp = 0.01)     9.9 87.522 < 2.2e-16 ***
## s(nday, sp = 0.01)     61.7 15.955 2.887e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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)
summary(CCC05.gam3)
```

```
##
## Call: gam(formula = Value ~ s(nyear, sp = 0.25) + s(nmonth, sp = 0.25) +
##       s(nday, sp = 0.25), data = CCC05)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36086 -0.20307  0.04743  0.20540  0.81026
##
## (Dispersion Parameter for gaussian family taken to be 0.2527)
##
```

```

##      Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 15.182 on 60.0689 degrees of freedom
## AIC: 219.5961
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 0.25)    1.000 11.2607 11.2607 44.5538 8.904e-09 ***
## s(nmonth, sp = 0.25)    1.000  0.2210  0.2210  0.8743  0.35351
## s(nday, sp = 0.25)      1.000  1.3014  1.3014  5.1492  0.02686 *
## Residuals              60.069 15.1820  0.2527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##              Npar Df  Npar F      Pr(F)
## (Intercept)
## s(nyear, sp = 0.25)      6.5  3.7308  0.002579 **
## s(nmonth, sp = 0.25)     7.3 18.7427 3.233e-13 ***
## s(nday, sp = 0.25)     45.2  3.9190 5.668e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

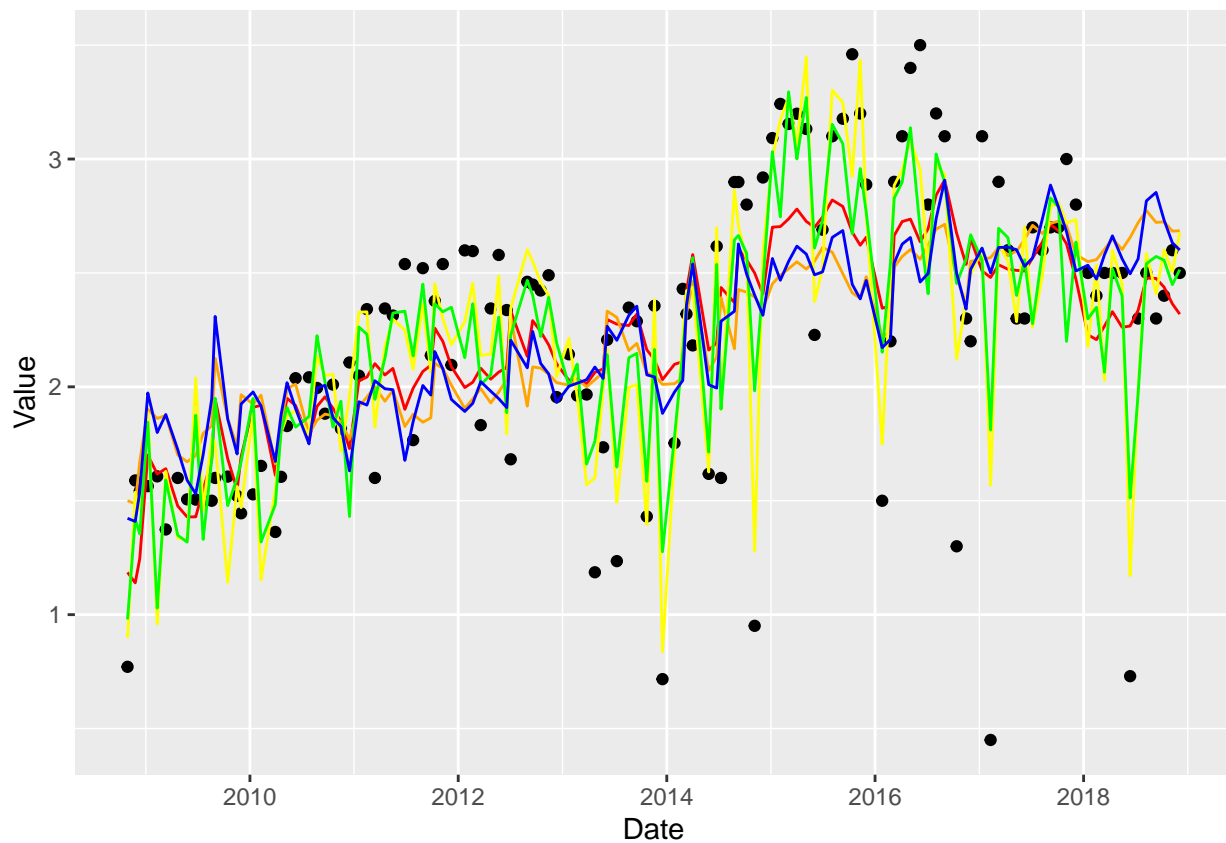
CCC05.gam4 <- gam(Value ~ s(nyear, sp=0.75) + s(nmonth, sp = 0.75) +
                  s(nday, sp = 0.75), data = CCC05)
pred4 <- predict(CCC05.gam4)
summary(CCC05.gam4)

##
## Call: gam(formula = Value ~ s(nyear, sp = 0.75) + s(nmonth, sp = 0.75) +
##      s(nday, sp = 0.75), data = CCC05)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.04990 -0.26239 -0.01294  0.35989  1.03976
##
## (Dispersion Parameter for gaussian family taken to be 0.2899)
##
##      Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 32.5685 on 112.3503 degrees of freedom
## AIC: 208.9111
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 0.75)    1.00 11.876 11.8764 40.9694 3.693e-09 ***
## s(nmonth, sp = 0.75)    1.00  0.126  0.1258  0.4339  0.511429
## s(nday, sp = 0.75)      1.00  2.015  2.0148  6.9502  0.009566 **
## Residuals             112.35 32.568  0.2899
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects

```

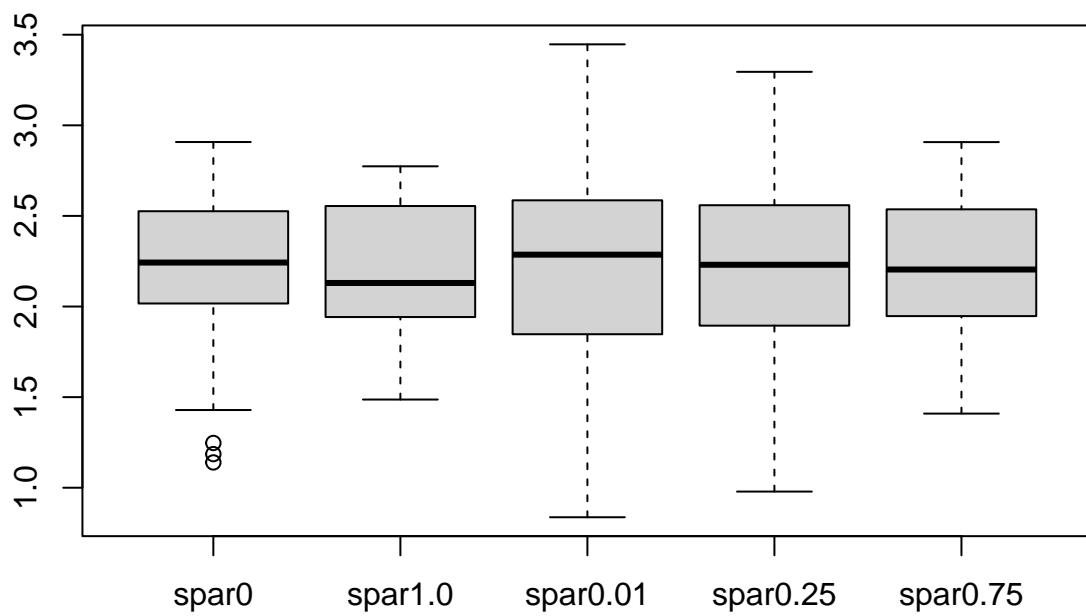
```
##              Npar Df  Npar F    Pr(F)
## (Intercept)
## s(nyear, sp = 0.75)      0.1 16.8892 0.01278 *
## s(nmonth, sp = 0.75)    0.2  0.0438 0.43071
## s(nday, sp = 0.75)      6.3  1.4286 0.20731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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")
```

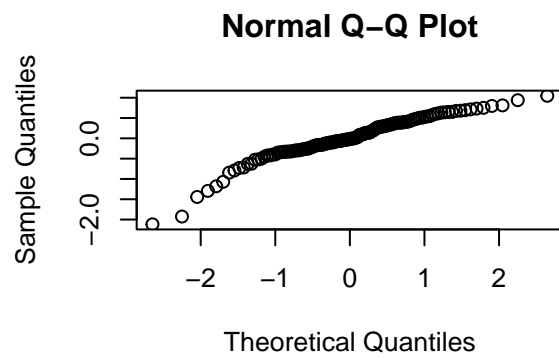
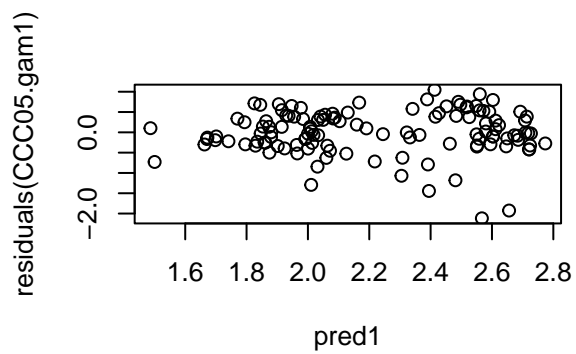
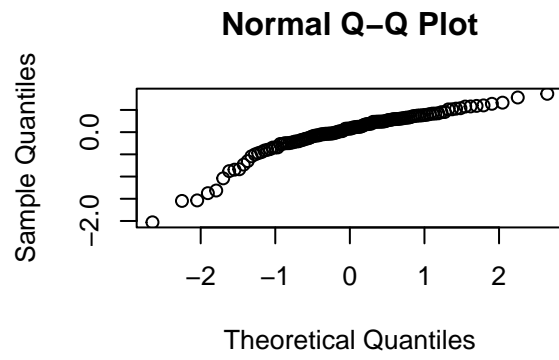
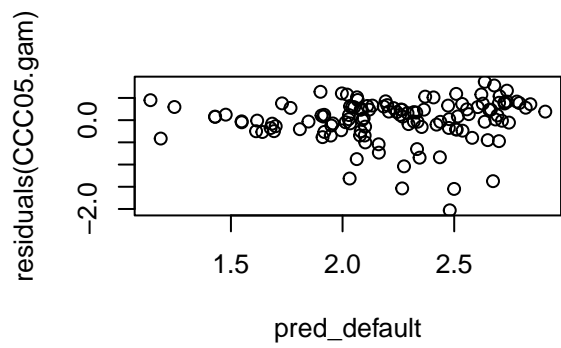


```
data <- data.frame(spar0 = fitted.values(CCC05.gam),
  spar1.0 = fitted.values(CCC05.gam1),
  spar0.01 = fitted.values(CCC05.gam2),
  spar0.25 = fitted.values(CCC05.gam3),
  spar0.75 = fitted.values(CCC05.gam4)
)
```

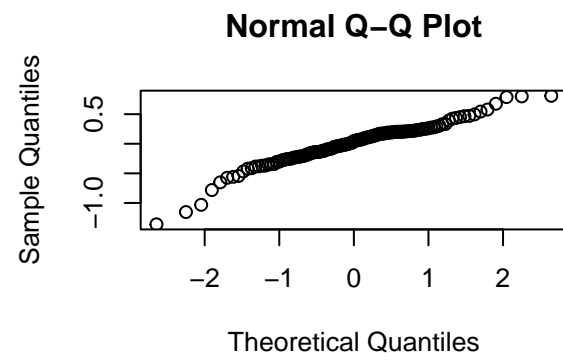
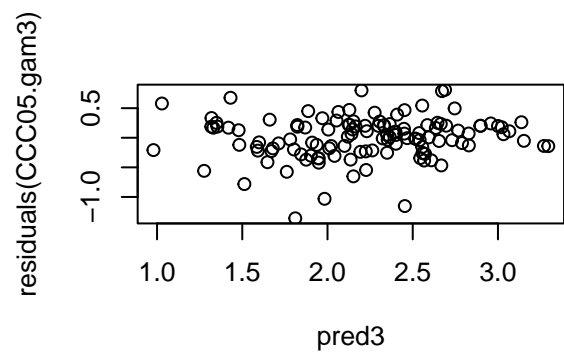
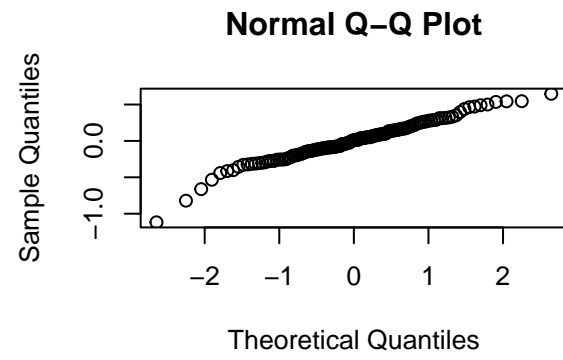
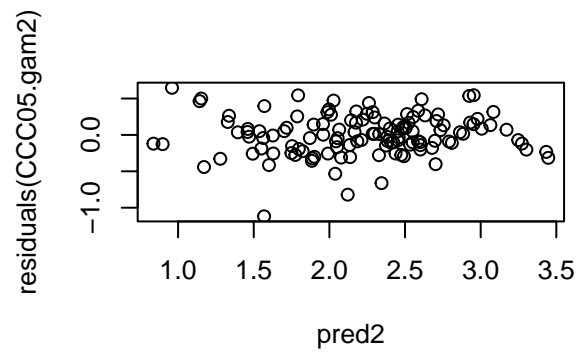
```
boxplot(data)
```



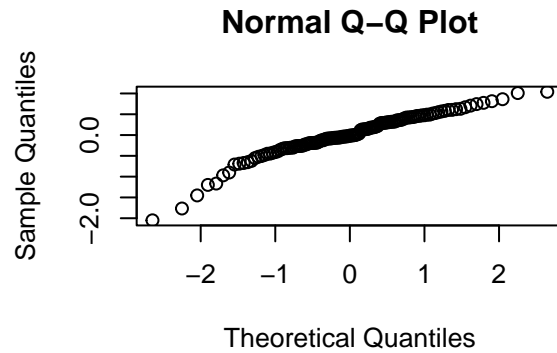
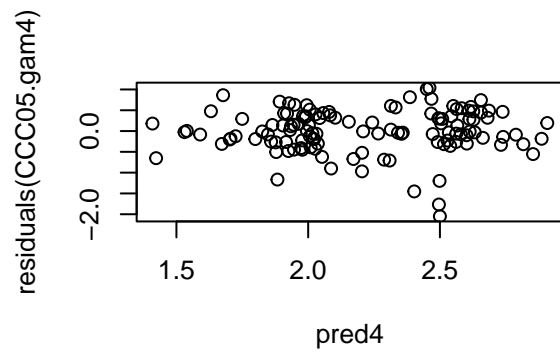
```
par(mfrow=c(2,2))
plot(pred_default, residuals(CCC05.gam),)
qqnorm(residuals(CCC05.gam))
plot(pred1, residuals(CCC05.gam1))
qqnorm(residuals(CCC05.gam1))
```



```
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

```
ECAN93.gam <- gam(Value ~ s(nyear) + s(nmonth) + s(nday), data = ECAN93 )
pred_default <- predict(ECAN93.gam)
summary(ECAN93.gam)
```

```
##
## Call: gam(formula = Value ~ s(nyear) + s(nmonth) + s(nday), data = ECAN93)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.61712 -0.24791  0.05168  0.30614  1.31908
##
## (Dispersion Parameter for gaussian family taken to be 0.4072)
##
##      Null Deviance: 271.865 on 178 degrees of freedom
## Residual Deviance: 67.6002 on 165.9998 degrees of freedom
## AIC: 361.6745
```

```
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##           Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear)   1 187.412 187.412 460.2108 < 2.2e-16 ***
## s(nmonth)  1  12.344  12.344  30.3131 1.374e-07 ***
## s(nday)    1   0.135   0.135   0.3325  0.565
## Residuals 166  67.600   0.407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##           Npar Df Npar F      Pr(F)
## (Intercept)
## s(nyear)           3 1.6923  0.17064
## s(nmonth)          3 9.1194 1.274e-05 ***
## s(nday)            3 3.5343  0.01612 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ECAN93.gam$y
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
## 4.1 4.2 4.2 4.0 4.4 4.6 4.4 4.5 4.5 4.2 4.3 4.2 4.4 4.3 4.2 4.4 4.3 4.6 4.6 4.6
## 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
## 4.4 4.4 4.3 4.5 4.1 4.0 4.3 4.1 4.4 4.4 4.9 5.2 5.8 5.8 5.3 5.5 5.9 5.2 5.1 5.1
## 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
## 5.3 5.8 5.3 5.4 4.9 5.1 4.7 4.5 4.3 4.2 4.7 4.9 5.0 5.4 5.3 6.0 6.1 6.2 5.6 5.5
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
## 5.4 5.5 5.6 5.8 6.1 6.0 5.8 5.9 5.9 5.7 5.7 5.2 5.3 5.1 5.3 5.5 5.7 5.2 5.8 4.5
## 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
## 6.8 4.5 6.5 3.9 6.9 7.1 6.9 6.5 6.3 6.5 6.0 6.6 6.6 6.4 5.9 3.0 5.8 6.2 5.6 6.3
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
## 6.4 1.7 6.3 5.9 5.8 5.5 6.6 6.1 5.6 6.0 6.1 6.4 6.6 6.7 7.4 7.9 7.8 7.2 6.8 6.4
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## 7.0 7.3 7.0 7.8 7.7 8.0 7.6 7.5 7.3 7.4 6.9 6.5 6.4 6.4 6.6 7.1 6.9 7.2 6.7 7.2
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
## 6.7 6.6 6.7 6.5 6.8 6.5 7.2 6.9 7.4 7.4 7.1 7.0 7.1 6.8 6.9 7.0 7.0 7.3 6.5 7.0
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
## 7.0 7.5 7.8 7.6 7.8 8.0 7.8 7.1 5.9 7.0 7.6 7.4 7.3 8.3 8.3 8.4 8.0 7.9 8.1
```

```
ECAN93.gam$X
```

```
## NULL
```

```
ECAN93.gam1 <- gam(Value ~ s(nyear, sp=1.0) + s(nmonth, sp = 1.0) +
                    s(nday, sp = 1.0), data = ECAN93)
pred1 <- predict(CCC05.gam1)
summary(ECAN93.gam1)
```

```
##
```

```
## Call: gam(formula = Value ~ s(nyear, sp = 1) + s(nmonth, sp = 1) +
##       s(nday, sp = 1), data = ECAN93)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.67518 -0.28111  0.03948  0.29924  1.38015
##
## (Dispersion Parameter for gaussian family taken to be 0.4135)
##
## Null Deviance: 271.865 on 178 degrees of freedom
## Residual Deviance: 71.5782 on 173.1225 degrees of freedom
## AIC: 357.6645
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 1)    1.00 185.391 185.391 448.3950 < 2.2e-16 ***
## s(nmonth, sp = 1)    1.00  12.375  12.375  29.9317 1.547e-07 ***
## s(nday, sp = 1)      1.00   0.129   0.129   0.3126  0.5768
## Residuals          173.12  71.578   0.413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##              Npar Df  Npar F    Pr(F)
## (Intercept)
## s(nyear, sp = 1)      0.0 0.27713 0.03056 *
## s(nmonth, sp = 1)     0.0 0.50310 0.01365 *
## s(nday, sp = 1)      1.9 3.01404 0.05536 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ECAN93.gam2 <- gam(Value ~ s(nyear, sp=0.01) + s(nmonth, sp = 0.01) +
                    s(nday, sp = 0.01), data = ECAN93)
pred2 <- predict(ECAN93.gam2)
summary(ECAN93.gam2)
```

```
##
## Call: gam(formula = Value ~ s(nyear, sp = 0.01) + s(nmonth, sp = 0.01) +
##       s(nday, sp = 0.01), data = ECAN93)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.735764 -0.241031 -0.003287  0.187134  1.588007
##
## (Dispersion Parameter for gaussian family taken to be 0.3015)
##
## Null Deviance: 271.865 on 178 degrees of freedom
## Residual Deviance: 25.3447 on 84.0508 degrees of freedom
## AIC: 349.9663
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## s(nyear, sp = 0.01) 1.000 169.259 169.259 561.3163 < 2.2e-16 ***
## s(nmonth, sp = 0.01) 1.000 12.600 12.600 41.7869 6.344e-09 ***
## s(nday, sp = 0.01) 1.000 0.029 0.029 0.0976 0.7555
## Residuals 84.051 25.345 0.302
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
## Npar Df Npar F Pr(F)
## (Intercept)
## s(nyear, sp = 0.01) 12.9 2.775 0.002577 **
## s(nmonth, sp = 0.01) 9.9 74.760 < 2.2e-16 ***
## s(nday, sp = 0.01) 68.2 12.293 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ECAN93.gam3 <- gam(Value ~ s(nyear, sp=0.25) + s(nmonth, sp = 0.25) +
                    s(nday, sp = 0.25), data = ECAN93)
pred3 <- predict(ECAN93.gam3)
summary(ECAN93.gam3)
```

```
##
## Call: gam(formula = Value ~ s(nyear, sp = 0.25) + s(nmonth, sp = 0.25) +
##          s(nday, sp = 0.25), data = ECAN93)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.85861 -0.22425 -0.01348  0.24210  1.14952
##
## (Dispersion Parameter for gaussian family taken to be 0.3098)
##
## Null Deviance: 271.865 on 178 degrees of freedom
## Residual Deviance: 34.067 on 109.9649 degrees of freedom
## AIC: 351.0789
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
## Df Sum Sq Mean Sq F value Pr(>F)
## s(nyear, sp = 0.25) 1.00 137.235 137.235 442.981 < 2.2e-16 ***
## s(nmonth, sp = 0.25) 1.00 13.659 13.659 44.089 1.237e-09 ***
## s(nday, sp = 0.25) 1.00 7.043 7.043 22.735 5.738e-06 ***
## Residuals 109.96 34.067 0.310
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
## Npar Df Npar F Pr(F)
## (Intercept)
## s(nyear, sp = 0.25) 9.4 2.9618 0.002989 **
## s(nmonth, sp = 0.25) 7.3 17.1267 1.776e-15 ***
## s(nday, sp = 0.25) 48.3 3.4062 5.741e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

ECAN93.gam4 <- gam(Value ~ s(nyear, sp=0.75) + s(nmonth, sp = 0.75) +
                  s(nday, sp = 0.75), data = ECAN93)
pred4 <- predict(ECAN93.gam4)
summary(ECAN93.gam4)

##
## Call: gam(formula = Value ~ s(nyear, sp = 0.75) + s(nmonth, sp = 0.75) +
##       s(nday, sp = 0.75), data = ECAN93)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.60281 -0.27861  0.03792  0.29949  1.26901
##
## (Dispersion Parameter for gaussian family taken to be 0.4149)
##
## Null Deviance: 271.865 on 178 degrees of freedom
## Residual Deviance: 69.4372 on 167.3419 degrees of freedom
## AIC: 363.7898
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## s(nyear, sp = 0.75)    1.00 186.155 186.155 448.630 < 2.2e-16 ***
## s(nmonth, sp = 0.75)    1.00  12.372  12.372  29.816 1.691e-07 ***
## s(nday, sp = 0.75)      1.00   0.126   0.126   0.304  0.5821
## Residuals             167.34  69.437   0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##              Npar Df  Npar F  Pr(F)
## (Intercept)
## s(nyear, sp = 0.75)    0.5 0.36689 0.3889
## s(nmonth, sp = 0.75)  0.2 0.29010 0.2919
## s(nday, sp = 0.75)    7.0 1.47183 0.1808

# ggplot(data = ECAN93, aes(x = Date, y = Value)) +
#   geom_point() +
#   geom_line(aes(x=Date, y=pred_default), colour = "pink") +
#   geom_line(aes(x=Date, y=pred1), colour = "lightblue") +
#   geom_line(aes(x=Date, y=pred2), colour = "lightgreen") +
#   geom_line(aes(x=Date, y=pred3), colour = "purple") +
#   geom_line(aes(x=Date, y=pred4), colour = "black")

```

QUESTION 2 Multiple Comparisons

```

# Read in dataset
library(xlsx)
herbicides <- read.xlsx("herbicides.xlsx", sheetIndex = 1)

```

- (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?