BIOS6643. L16 Introduction to Generalized Additive Models (GAMs)

Modeling non-linearity in longitudinal data

Colorado CF FEV data

The data set gives characteristics of children patients with a diagnosis of Cystic Fibrosis (CF) who are patients at the Colorado Children's Hospital.

VARIABLE DESCRIPTIONS:
id: patient ID
race: race
age: Age in years
gender : gender fev: % of predicted forced expiratory volume in 1 second

```
#  Read in the data

##  Read in the data
dat <- read.csv("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/BIOS664
                header=TRUE)
length(unique(dat$id)) ## number of individuals
```

```
## [1] 723
```

```
head(dat,3)
```

```
##   id      Race Genotypes1 Genotypes2   age gender      fev
## 1  1 Caucasian       F508      R553X 10.32      m 55.45775
## 2  1 Caucasian       F508      R553X 10.47      m 53.48924
## 3  1 Caucasian       F508      R553X 10.72      m 49.62416
```

```
summary(dat$age) ## age in years
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.01    9.83   13.50   13.73   17.29   25.00
```

```
summary(dat$fev)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.99   66.72   85.53   81.90   99.31  160.10
```
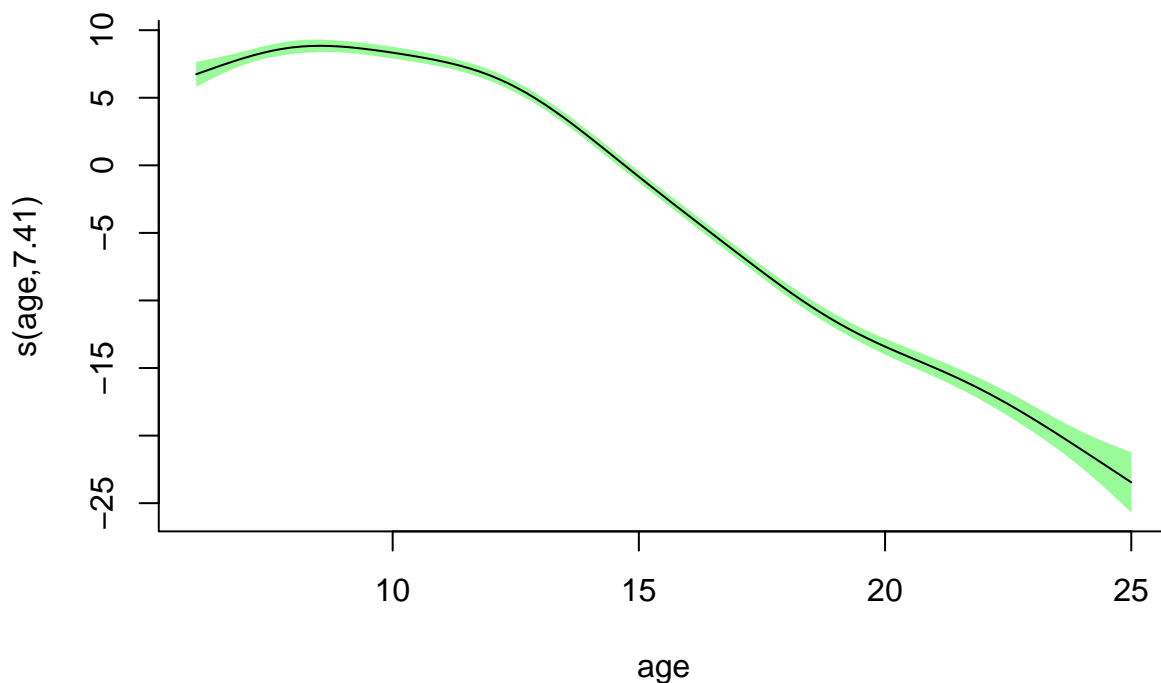
```
dat$sexf <- ifelse(dat$gender=='f', 1, 0)
table(dat$gender, dat$sex)
```

```
##
##       0    1
##   f    0 9716
##   m 8778    0
```
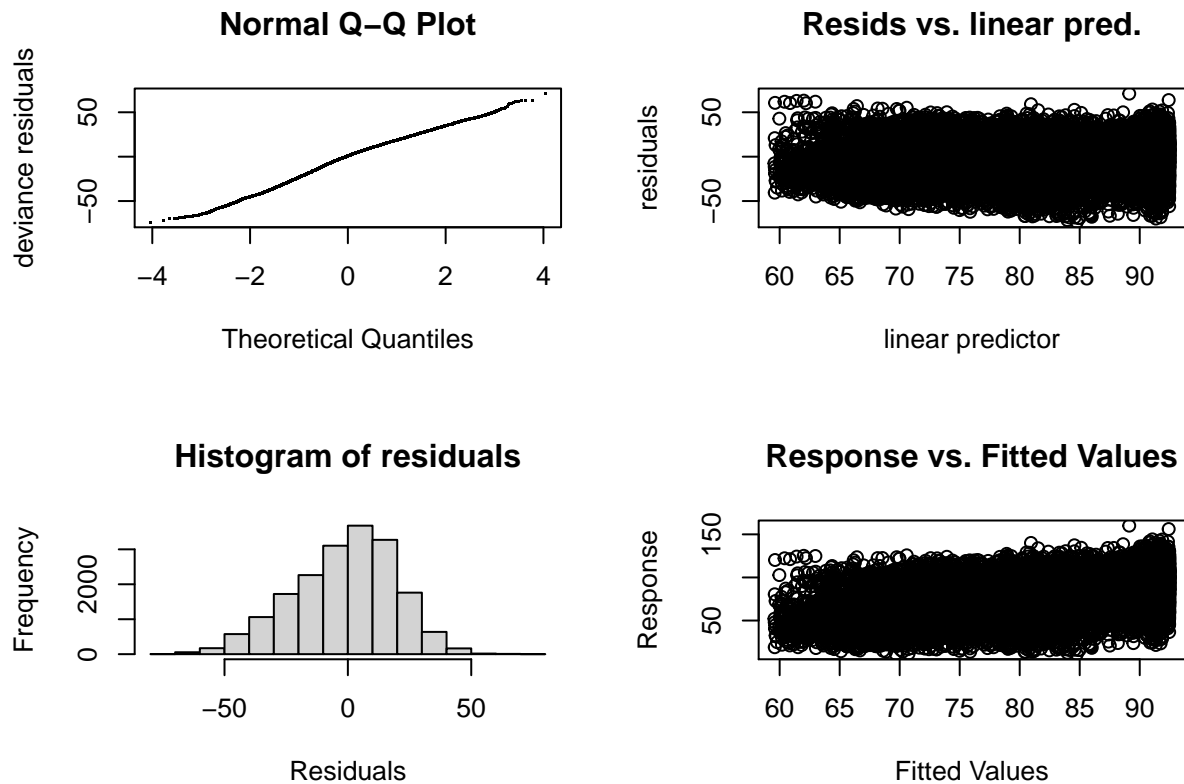
Smooth effect of age

```
fit <- gamm(fev ~ sexf + s(age) ,
            random = list(id = ~1), method='REML',
            data = dat)
##attributes(fit)
summary(fit$gam)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## fev ~ sexf + s(age)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.6174     0.9418  88.787   <2e-16 ***
## sexf         -0.5895     1.3619  -0.433    0.665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df     F p-value
## s(age) 7.413  7.413 686.9  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.233
##   Scale est. = 128.72    n = 18494
```

```
plot(fit$gam, shade=TRUE, shade.col='palegreen', bty='l')
```

```
gam.check(fit$gam)
```


Normal Q–Q Plot — *deviance residuals* vs *Theoretical Quantiles*

Resids vs. linear pred. — *residuals* vs *linear predictor*

Histogram of residuals — *Frequency* vs *Residuals*

Response vs. Fitted Values — *Response* vs *Fitted Values*

```
##
## 'gamm' based fit - care required with interpretation.
## Checks based on working residuals may be misleading.
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##         k'  edf k-index p-value
## s(age) 9.00 7.41    0.95  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit$lme)
```

```
## Linear mixed-effects model fit by REML
##   Data: strip.offset(mf)
##      AIC      BIC   logLik
##   144961 145007.9 -72474.5
##
## Random effects:
##  Formula: ~Xr - 1 | g
##  Structure: pdIdnot
##                 Xr1       Xr2       Xr3       Xr4       Xr5       Xr6       Xr7       Xr8
## StdDev: 13.23686 13.23686 13.23686 13.23686 13.23686 13.23686 13.23686 13.23686
##
##  Formula: ~1 | id %in% g
##         (Intercept) Residual
## StdDev:    17.79025 11.34559
```

```
## 
## Fixed effects:  y ~ X - 1
##                 Value Std.Error    DF  t-value p-value
## X(Intercept) 83.61741 0.9417719 17770 88.78733  0.0000
## Xsexf        -0.58949 1.3618561   721 -0.43286  0.6652
## Xs(age)Fx1   -2.61644 1.7107117 17770 -1.52945  0.1262
##  Correlation:
##            X(Int) Xsexf
## Xsexf      -0.691
## Xs(age)Fx1 -0.007  0.002
## 
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -7.49278498 -0.54586527  0.04401691  0.60276221  5.74343107
## 
## Number of Observations: 18494
## Number of Groups:
##       g id %in% g
##       1       723
```

Weight loss trial (PI: Bessesen)

- Randomized trial at the subject level to compare a toolboox intervention versus usual care in the primary care setting (Saxon, 2019 J Gen Intern Med)

- The toolbox consisted of: partial meal replacement program, Weight Watchers vouchers, recreation center membership, phentermine-topiramate ER, phentermine, or a group behavioral weight loss program (Colorado Weigh).

- 305 individuals were randomly selected to be offered intervention, and 2640 were eligible comparators in the usual care group

- 119 individuals had a baseline visit (305-119=186 did not consent or did not attend baseline visit)

Fit a GAMM with a smooth effect of time (timemo) for the weight loss trial. Adjust for treatment arm (studygrp)

```
dat.wt <- read.csv("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/BIOS
head(dat.wt, 2)
```

```
##   sex age ethnic language  bmi  studyid diabetes osteo_arthritis back_pain
## 1   2  35      1        2 34.5 10020502        1               0         0
## 2   2  35      1        2 34.5 10020502        1               0         0
##   pre_diabetes hypertension hyperlipidemia metabolic_syndrome
## 1            0            0              1                  0
## 2            0            0              1                  0
##   cerebrovascular_disease sleep_disorder peripheral_vascular_disease
## 1                       0              0                           0
## 2                       0              0                           0
##   congestive_heart_failure racecat num_comorbs studygrp agecat raceth cadath
## 1                        0       5           2        2      3      4      0
## 2                        0       5           2        2      3      4      0
##   diabhyp bmicat    X_LABEL_     wtkg   wtchgkg timemo site
```

```
## 1        1        4 Weight (kg) 77.28125  0.000000    0.00    6
## 2        1        4                75.90852 -1.372727    4.08    6
```