Answers to exercises Session 5

Marjolein Fokkema

Exercise 1

Read in data:

```
load("MASQ.Rda")
set.seed(1)
train <- sample(1:nrow(MASQ), size = nrow(MASQ)*.8)
summary(MASQ)
```

```
##  D_DEPDYS       AD               AA              GDD             GDA
##  0:1927   Min.   : 26.00   Min.   :17.00   Min.   :12.00   Min.   :11.0
##  1:1670   1st Qu.: 64.00   1st Qu.:22.00   1st Qu.:20.00   1st Qu.:19.0
##           Median : 77.00   Median :28.00   Median :29.00   Median :24.0
##           Mean   : 75.05   Mean   :32.01   Mean   :30.64   Mean   :25.4
##           3rd Qu.: 88.00   3rd Qu.:39.00   3rd Qu.:40.00   3rd Qu.:31.0
##           Max.   :110.00   Max.   :83.00   Max.   :60.00   Max.   :54.0
##       GDM          leeftijd    geslacht
##  Min.   :15.0   Min.   :17.0   m:1317
##  1st Qu.:31.0   1st Qu.:28.0   v:2280
##  Median :40.0   Median :38.0
##  Mean   :40.6   Mean   :38.8
##  3rd Qu.:50.0   3rd Qu.:48.0
##  Max.   :75.0   Max.   :91.0
```

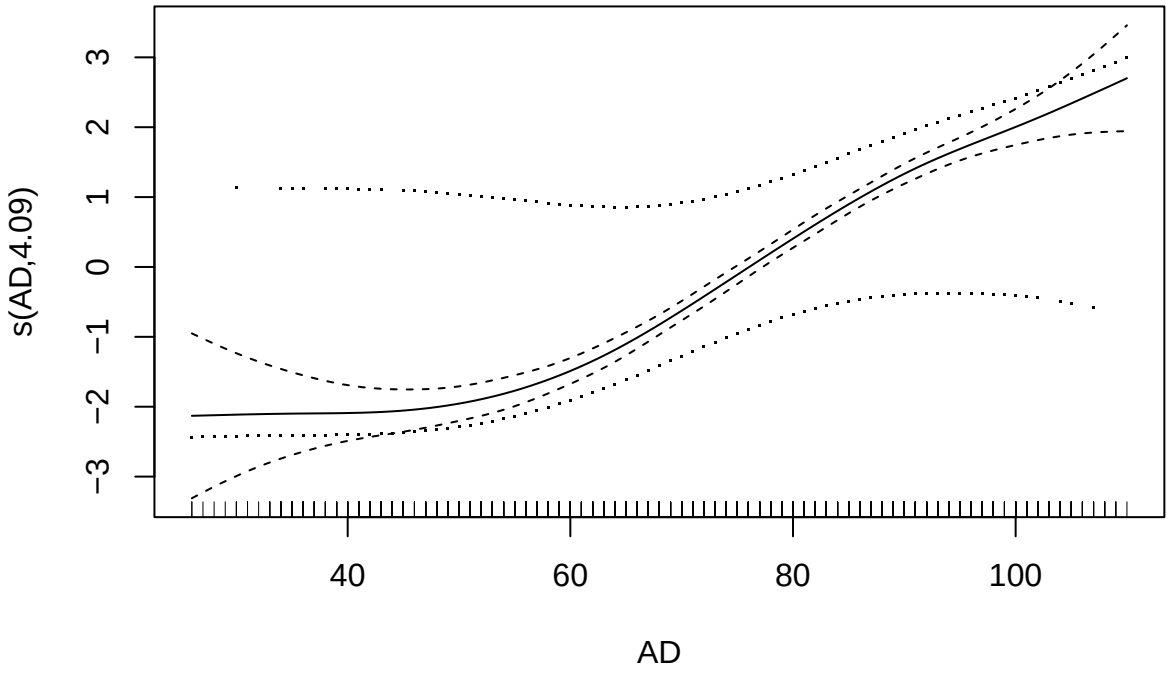
Fit a smoothing spline of the `AD` variable to predict `D_DEPDYS`:

```
library("mgcv")
GAM <- gam(D_DEPDYS ~ s(AD, bs = "cr"),
           data = MASQ[train, ], method = "REML", family = "binomial")
summary(GAM)
```

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## D_DEPDYS ~ s(AD, bs = "cr")
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.24089    0.04751   -5.07 3.98e-07 ***
## ---
```

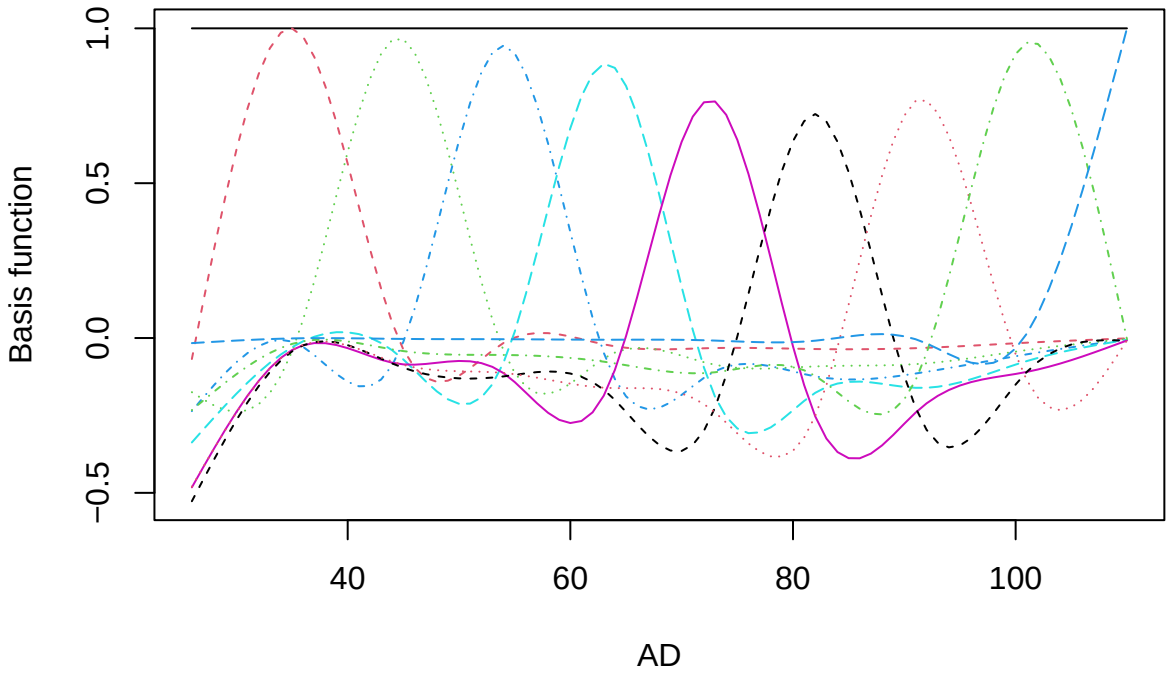
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##        edf Ref.df Chi.sq p-value
## s(AD) 4.09  5.041  672.4  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.297   Deviance explained = 23.7%
## -REML = 1522.2  Scale est. = 1         n = 2877
```

```
plot(GAM, residuals = TRUE)
```



Inspect the basis functions that were created for `AD`:

```
mod_mat <- model.matrix(GAM)
matplot(sort(MASQ$AD[train]), mod_mat[order(MASQ$AD[train]), ], type = "l",
        xlab = "AD", ylab = "Basis function")
```


Basis function
1.0
0.5
0.0
−0.5
40
60
80
100
AD