Exposure to Dust - Trees

January 25, 2024

First the dust data are loaded from the package "catdata".

```
library(catdata)
data(dust)
```

Trees can be fitted by use of the function "rpart" from package "rpart".

```
library(rpart)
```

Now a tree is fitted. We take "years" as the only covariate, "bronch" is the binary response. Afterwards the corresponding tree is plotted.

```
tree1 <-rpart(as.factor(bronch) ~ years, data = dust,   method = "class",
              control = rpart.control(cp = 0.001,  parms=list(split='information'),
                                      maxdepth = 4))
plot(tree1, xpd=TRUE)
text(tree1)
```

In the following the fit is plotted. It shows how the tree can be interpreted as regression function.

```
        pred <- predict(tree1)
        year<- dust$years
        year [dust$years<15.5] <- 1
        year [dust$years>15.5 & dust$years<36.5] <- 2
        year [dust$years>36.5 & dust$years<47.5] <- 3
        year [dust$years>47.5 & dust$years<50.5] <- 4
        year [dust$years>50.5] <- 5

pre5 <- unique( pred[,2][year==5])
pre4 <- unique( pred[,2][year==4])
pre3 <- unique( pred[,2][year==3])
pre2 <- unique( pred[,2][year==2])
pre1 <- unique( pred[,2][year==1])

meanyear <- c()

for (i in min(dust$years):max(dust$years)){
meanyear[i] <- sum(dust$bronch[dust$year==i])
```

```
if(sum(dust$bronch[dust$year==i])!=0){
meanyear[i] <- mean(dust$bronch[dust$year==i])
}
}

dust$means<- rep(2, nrow(dust))

 for (k in 1:nrow(dust)){
 dust$means[k] <- meanyear[dust$years[k]]
 }
```

```
plot(dust$years, dust$means, xlab="years",ylab="")
segments(x0=3,x1=15.5,y0=pre1)
segments(x0=15.5,x1=15.5,y0=pre1,y1=pre2)
segments(x0=15.5,x1=36.5,y0=pre2)
segments(x0=36.5,x1=36.5,y0=pre2,y1=pre3)
segments(x0=36.5,x1=47.5,y0=pre3)
  segments(x0=47.5,x1=47.5,y0=pre3,y1=pre4)
segments(x0=47.5,x1=50.5,y0=pre4)
  segments(x0=50.5,x1=50.5,y0=pre4,y1=pre5)
segments(x0=50.5,x1=66,y0=pre5)
```

An alternative package to generate trees is "party" which contains the function "ctree".

```
library(party)
```

As before with "rpart" we fit a tree with "years" as only covariate.

```
treeP1 <-ctree(as.factor(bronch) ~ years, data = dust)
plot(treeP1)
```

```
        year<- dust$years
        year [dust$years<7.5] <- 1
        year [dust$years>7.5 & dust$years<15.5] <- 2
        year [dust$years>15.5 & dust$years<36.5] <- 3
        year [dust$years>36.5] <- 4

pre4 <- mean(dust$bronch[year==4])
pre3 <- mean(dust$bronch[year==3])
pre2 <- mean(dust$bronch[year==2])
pre1 <- mean(dust$bronch[year==1])
```

```
plot(dust$years, dust$means, xlab="years",ylab="")
segments(x0=3,x1=7.5,y0=pre1)
segments(x0=7.5,x1=7.5,y0=pre1,y1=pre2)
```

```
segments(x0=7.5,x1=15.5,y0=pre2)
segments(x0=15.5,x1=15.5,y0=pre2,y1=pre3)
segments(x0=15.5,x1=36.5,y0=pre3)
 segments(x0=36.5,x1=36.5,y0=pre3,y1=pre4)
segments(x0=36.5,x1=66,y0=pre4)
```

Now we take "smoke", "years" and "dust" as covariates for the binary response "bronch" and again plot the tree.

```
treeP2 <-ctree(as.factor(bronch) ~ smoke + years + dust, data = dust)
plot(treeP2)
```