

Duration of Unemployment - Trees

January 25, 2024

```
library(catdata)
data(unemployment, package="catdata")
```

To fit a tree for the unemployment data we use "ctree" from the library "party".

```
library(party)
tree1<-ctree(as.factor(durbin)~age,data=unemployment)

plot(tree1)
```

The fitted regression function can be obtained by computing the respective means within the identified regions and plot them as function of age.

```
unemployment$durbin[unemployment$durbin==2]<-0
year<- unemployment$age
year [unemployment$age<29.5] <- 1
year [unemployment$age>29.5 & unemployment$age<52.5] <- 2
year [unemployment$age>52.5] <- 3

pre3 <- mean(unemployment$durbin[year==3])
pre2 <- mean(unemployment$durbin[year==2])
pre1 <- mean(unemployment$durbin[year==1])

meanyear <- c()

for (i in min(unemployment$age):max(unemployment$age)){
  meanyear[i] <- sum(unemployment$durbin[unemployment$age==i])
  if(sum(unemployment$durbin[unemployment$age==i])!=0){
    meanyear[i] <- mean(unemployment$durbin[unemployment$age==i])
  }
}

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

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

```
plot(unemployment$age, unemployment$means, xlab="age", ylab="", cex.axis=1.5,
      cex.lab=1.5)
segments(x0=min(unemployment$age), x1=29.5, y0=pre1)
segments(x0=29.5, x1=29.5, y0=pre1, y1=pre2)
segments(x0=29.5, x1=52.5, y0=pre2)
segments(x0=52.5, x1=52.5, y0=pre2, y1=pre3)
segments(x0=52.5, x1=max(unemployment$age), y0=pre3)
```