DanielPonikowski_PD1

Daniel Ponikowski

5 marca 2019

```
options(stringsAsFactors = FALSE)
library(readstata13)
library(dplyr)
library(caret)
library(ggplot2)
data <- read.dta13(file = "HCMST 2017 fresh sample for public sharing draft v1.1.dta")
```

Wybrane zmienne :

+ ppwork - aktualny status zatrudnienia
+ w6_q20 - czy obecnie mieszkasz z partnerem?
+ Q21A_Year - w ktorym roku pierwszy raz spotkałes partnera?
+ ppage - wiek

```
df <- data[,c("S1","ppwork","w6_q19","Q21A_Year","ppage")]


df <- df %>% mutate(Q21A_Year = as.numeric(as.character(Q21A_Year))
                    ,ppwork = factor(ppwork)
                    ,w6_q19 = factor(w6_q19)
                    ,ppage = as.numeric(ppage)
                    ,S1= factor(S1)) %>%
  na.omit() %>% unique() %>% as.data.frame()

control <- trainControl(method = "cv", number=10, search = "random")
metric <- "Accuracy"
RF <- train(df[2:5],df$S1 ,method = "rf", metric = metric,
                  trControl = control)
```

Reczne narysowanie wykresu ceteris paribus dla losowej osoby.

Najpierw wylosuje osobe dla ktorej bedziemy rysowac wykresy.

```
(os <- df[sample(1:nrow(df),size = 1),])
```

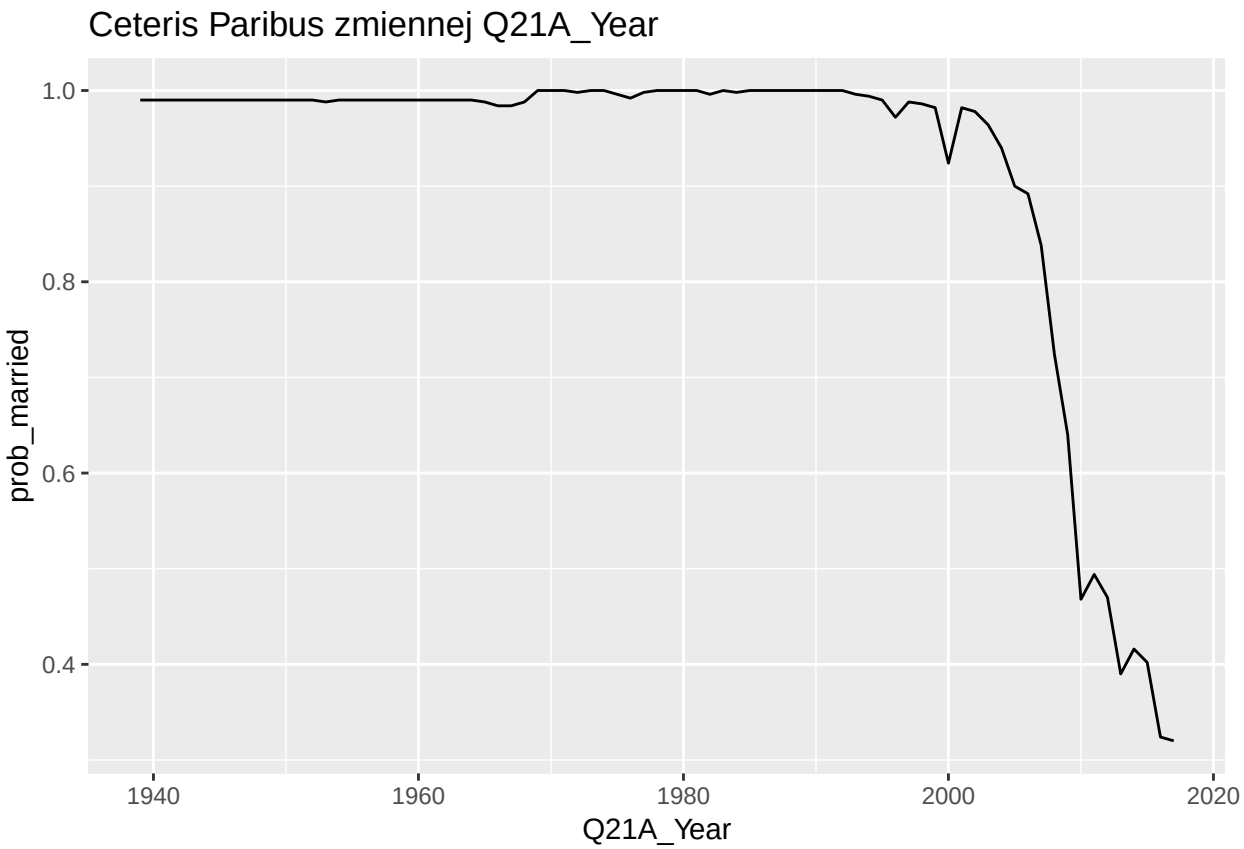
```
##                      S1                        ppwork w6_q19 Q21A_Year ppage
## 2957 Yes, I am Married Working - as a paid employee    Yes      2000    51
```

Zmienna Q21A_Year

```
year <- min(df$Q21A_Year):max(df$Q21A_Year)
df_Q21A_Year <- data.frame(ppwork = rep(os$ppwork,length(year))
                           ,w6_q19 = rep(os$w6_q19,length(year))
                           ,Q21A_Year = year
                           ,ppage = rep(os$ppage,length(year)) )

df_Q21A_Year$prob_married <- predict(RF,df_Q21A_Year,type = "prob")[,1]
```

```
ggplot(df_Q21A_Year,aes(x = Q21A_Year,y = prob_married)) + geom_line() +
  ggtitle("Ceteris Paribus zmiennej Q21A_Year")
```

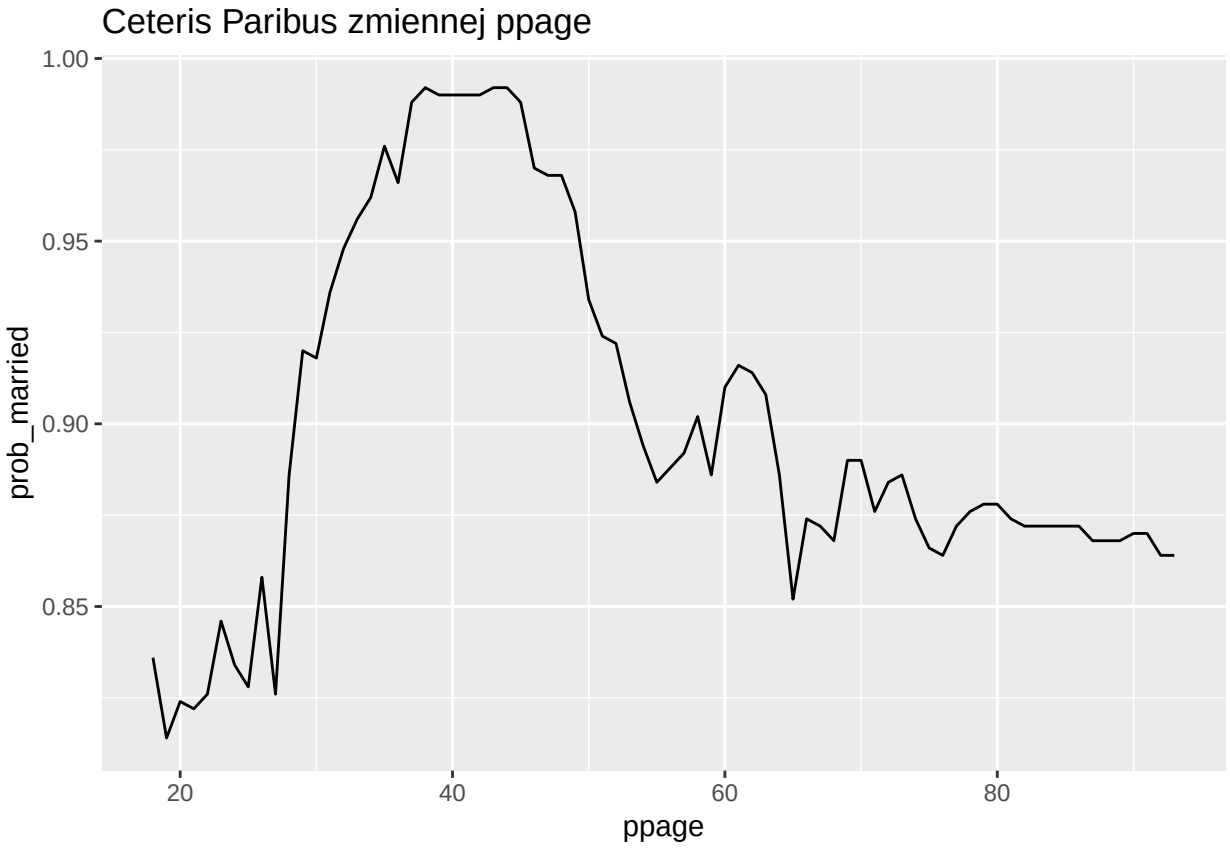


Zmienna ppage

```
age <- min(df$ppage):max(df$ppage)
df_ppage <- data.frame(ppwork = rep(os$ppwork,length(age))
                       ,w6_q19 = rep(os$w6_q19,length(age))
                       ,Q21A_Year = rep(os$Q21A_Year,length(age))
                       ,ppage = age )

df_ppage$prob_married <- predict(RF,df_ppage,type = "prob")[,1]

ggplot(df_ppage,aes(x = ppage,y = prob_married)) + geom_line() +
  ggtitle("Ceteris Paribus zmiennej ppage")
```


Ceteris Paribus zmiennej ppage
1.00
0.95
0.90
0.85
prob_married
20
40
60
80
ppage