Modelo\_inicial

Juan Ramon Lacalle

14/12/2018

## Preparación de datos

Incluimos el archivo original con los datos de trabajadores.

bajas <- read\_excel("Base\_TMLE\_transposta.xlsx",   
col\_types = c("text", "numeric", "numeric", "skip", "numeric", "numeric", "numeric", "text", "numeric", "numeric", "numeric", "text", "numeric", "numeric", "numeric"))

Y creamos una variable, **cambio**, que compara el código CIE de la baja, e indica si ha cambiado (1) con respecto a la visita anterior, o no (0).

bajas\_2 <- bajas %>%   
 group\_by(Num\_Unico) %>%   
 mutate(cambio = CIE - lag(CIE, default = CIE[1]))  
bajas\_2$cambio[1] <- 0  
bajas\_2$cambio[bajas\_2$cambio != 0] <- 1

A partir de ese fichero, definimos una nueva variable para codificar los cambios que se producen entre visitas.Es decir, si los cambios de la CIE pasan de una categoría a otra, según el siguiente esquema:

* 1: Cualquier código (excepto 5) -> Código 5.
* 2: Código 5 -> Cualquier código (excepto 5).
* 3: Cualquier código (excepto 5) -> Cualquier código (excepto 5).

bajas$g\_CIE <- recode(bajas$CIE, `5`=1, .default = 0)  
bajas\_4 <- bajas %>%   
 group\_by(Num\_Unico) %>%   
 mutate(cambio\_CIE = g\_CIE - lag(g\_CIE, default = g\_CIE[1]))  
bajas\_4$cambio\_CIE[bajas\_4$cambio\_CIE == 0] <- 0  
bajas\_4$cambio\_CIE[bajas\_4$cambio\_CIE == -1] <- 2  
bajas\_4$cambio\_CIE[bajas\_4$cambio\_CIE == 0 & bajas\_4$g\_CIE == 0] <- 3

## Estimación modelo 1

En el modelo inicial, usamos los algoritmos por defecto de Superlearner incluidos en tmle.

modelo1 <- tmle(Y=bajas\_2$READAPTADO,A=bajas\_2$cambio,W=bajas\_2[,c(7,8,9,10,14)], id=bajas\_2$Num\_Unico,  
 family="binomial")  
summary(modelo1)

## Initial estimation of Q  
## Procedure: SuperLearner  
## Model:  
## Y ~ SL.glm\_All + SL.step\_All + SL.glm.interaction\_All  
##   
## Coefficients:   
## SL.glm\_All 0.8306685   
## SL.step\_All 0   
## SL.glm.interaction\_All 0.1693315   
##   
## Estimation of g (treatment mechanism)  
## Procedure: SuperLearner   
## Model:  
## A ~ SL.glm\_All + SL.step\_All + SL.glm.interaction\_All   
##   
## Coefficients:   
## SL.glm\_All 0   
## SL.step\_All 0.05710754   
## SL.glm.interaction\_All 0.9428925   
##   
## Estimation of g.Z (intermediate variable assignment mechanism)  
## Procedure: No intermediate variable   
##   
## Estimation of g.Delta (missingness mechanism)  
## Procedure: No missingness   
##   
## Bounds on g: ( 0.025 0.975 )  
##   
## Additive Effect  
## Parameter Estimate: -0.028691  
## Estimated Variance: 0.00044087  
## p-value: 0.1718  
## 95% Conf Interval: (-0.069845, 0.012463)   
##   
## Additive Effect among the Treated  
## Parameter Estimate: -0.030457  
## Estimated Variance: 0.00032215  
## p-value: 0.089718  
## 95% Conf Interval: (-0.065636, 0.0047225)   
##   
## Additive Effect among the Controls  
## Parameter Estimate: -0.027611  
## Estimated Variance: 0.0006031  
## p-value: 0.26088  
## 95% Conf Interval: (-0.075745, 0.020523)   
##   
## Relative Risk  
## Parameter Estimate: 0.94155  
## p-value: 0.17669  
## 95% Conf Interval: (0.86278, 1.0275)   
##   
## log(RR): -0.060226  
## variance(log(RR)): 0.0019872   
##   
## Odds Ratio  
## Parameter Estimate: 0.89132  
## p-value: 0.17241  
## 95% Conf Interval: (0.75555, 1.0515)   
##   
## log(OR): -0.11505  
## variance(log(OR)): 0.0071092

#print(modelo2)

## Estimación modelo 2

Vamos a introducir el tipo de cambio de diagnóstico. Como tiene que ser dicotómica, creamos una nueva variable que recoge si ha tenido una baja por salud mental durante el seguimiento.

* nivel “1”=cualquier CIE salud mental.
* nivel “0”=cualquier CIE distinto salud mental.

bajas\_4$smental[bajas\_4$cambio\_CIE==0|bajas\_4$cambio\_CIE==3] <- 0

## Warning: Unknown or uninitialised column: 'smental'.

bajas\_4$smental[bajas\_4$cambio\_CIE==1|bajas\_4$cambio\_CIE==2] <- 1  
  
modelo2 <- tmle(Y=bajas\_4$READAPTADO,A=bajas\_4$smental,W=bajas\_4[,c(7,8,9,10)], id=bajas\_4$Num\_Unico,  
 family="binomial")  
summary(modelo2)

## Initial estimation of Q  
## Procedure: SuperLearner  
## Model:  
## Y ~ SL.glm\_All + SL.step\_All + SL.glm.interaction\_All  
##   
## Coefficients:   
## SL.glm\_All 0.5193145   
## SL.step\_All 0   
## SL.glm.interaction\_All 0.4806855   
##   
## Estimation of g (treatment mechanism)  
## Procedure: SuperLearner   
## Model:  
## A ~ SL.glm\_All + SL.step\_All + SL.glm.interaction\_All   
##   
## Coefficients:   
## SL.glm\_All 1   
## SL.step\_All 0   
## SL.glm.interaction\_All 0   
##   
## Estimation of g.Z (intermediate variable assignment mechanism)  
## Procedure: No intermediate variable   
##   
## Estimation of g.Delta (missingness mechanism)  
## Procedure: No missingness   
##   
## Bounds on g: ( 0.025 0.975 )  
##   
## Additive Effect  
## Parameter Estimate: 0.067576  
## Estimated Variance: 0.00078538  
## p-value: 0.015895  
## 95% Conf Interval: (0.012648, 0.1225)   
##   
## Additive Effect among the Treated  
## Parameter Estimate: 0.062936  
## Estimated Variance: 0.00068738  
## p-value: 0.016374  
## 95% Conf Interval: (0.011548, 0.11432)   
##   
## Additive Effect among the Controls  
## Parameter Estimate: 0.067998  
## Estimated Variance: 0.00080007  
## p-value: 0.016218  
## 95% Conf Interval: (0.012558, 0.12344)   
##   
## Relative Risk  
## Parameter Estimate: 1.143  
## p-value: 0.011601  
## 95% Conf Interval: (1.0303, 1.2681)   
##   
## log(RR): 0.13368  
## variance(log(RR)): 0.0028049   
##   
## Odds Ratio  
## Parameter Estimate: 1.311  
## p-value: 0.016358  
## 95% Conf Interval: (1.051, 1.6353)   
##   
## log(OR): 0.27076  
## variance(log(OR)): 0.012719

#print(modelo3)

## Modelo 3

En este modelo incluimos varios algoritmos de SuperLearner.

SL.library.2 <- c("SL.glm", "SL.step","SL.step.interaction","SL.glm.interaction","SL.gam","SL.randomForest","SL.rpart")  
modelo3 <- tmle(Y=bajas\_4$READAPTADO,A=bajas\_4$smental,W=bajas\_4[,c(7,8,9,10)], id=bajas\_4$Num\_Unico,  
 family="binomial", g.SL.library = SL.library.2)

## Loading required package: gam

## Loading required package: splines

## Loading required package: foreach

##   
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':  
##   
## accumulate, when

## Loaded gam 1.16

## Loading required package: randomForest

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

## The following object is masked from 'package:dplyr':  
##   
## combine

## Loading required package: rpart

summary(modelo3)

## Initial estimation of Q  
## Procedure: SuperLearner  
## Model:  
## Y ~ SL.glm\_All + SL.step\_All + SL.glm.interaction\_All  
##   
## Coefficients:   
## SL.glm\_All 0   
## SL.step\_All 0.8369439   
## SL.glm.interaction\_All 0.1630561   
##   
## Estimation of g (treatment mechanism)  
## Procedure: SuperLearner   
## Model:  
## A ~ SL.glm\_All + SL.step\_All + SL.step.interaction\_All + SL.glm.interaction\_All + SL.gam\_All + SL.randomForest\_All + SL.rpart\_All   
##   
## Coefficients:   
## SL.glm\_All 0   
## SL.step\_All 0   
## SL.step.interaction\_All 0   
## SL.glm.interaction\_All 0   
## SL.gam\_All 0.4871021   
## SL.randomForest\_All 0.1816585   
## SL.rpart\_All 0.3312394   
##   
## Estimation of g.Z (intermediate variable assignment mechanism)  
## Procedure: No intermediate variable   
##   
## Estimation of g.Delta (missingness mechanism)  
## Procedure: No missingness   
##   
## Bounds on g: ( 0.025 0.975 )  
##   
## Additive Effect  
## Parameter Estimate: 0.05845  
## Estimated Variance: 0.00054378  
## p-value: 0.012193  
## 95% Conf Interval: (0.012744, 0.10416)   
##   
## Additive Effect among the Treated  
## Parameter Estimate: 0.052144  
## Estimated Variance: 0.00059393  
## p-value: 0.032386  
## 95% Conf Interval: (0.0043773, 0.09991)   
##   
## Additive Effect among the Controls  
## Parameter Estimate: 0.059436  
## Estimated Variance: 0.00054267  
## p-value: 0.010729  
## 95% Conf Interval: (0.013777, 0.10509)   
##   
## Relative Risk  
## Parameter Estimate: 1.1236  
## p-value: 0.0094355  
## 95% Conf Interval: (1.029, 1.2269)   
##   
## log(RR): 0.11653  
## variance(log(RR)): 0.0020152   
##   
## Odds Ratio  
## Parameter Estimate: 1.2637  
## p-value: 0.012422  
## 95% Conf Interval: (1.0519, 1.5183)   
##   
## log(OR): 0.23407  
## variance(log(OR)): 0.0087667

# print(modelo3)