

# 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",  
bajas$Sexo <- as.factor(bajas$Sexo)
```

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$g_CIE == 0] <- 3  
bajas_4$smental <- 0  
bajas_4$smental[bajas_4$cambio_CIE==0|bajas_4$cambio_CIE==3] <- 0  
bajas_4$smental[bajas_4$cambio_CIE==1|bajas_4$cambio_CIE==2] <- 1
```

## 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")

## Loading required namespace: dbarts
## Loading required namespace: gam
## 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 1.16.1
prueba <- bajas_2[1:500,]
modelo1 <- tmle(Y=prueba$READAPTADO,
               A=prueba$cambio,
               W=prueba[,c(7,8,9,10,14)],
               id=prueba$Num_Unico,family="binomial")
summary(modelo1)

## Initial estimation of Q
## Procedure: cv-SuperLearner, ensemble
## Model:
##   Y ~ SL.glm_All + tmle.SL.dbarts2_All + SL.glmnet_All
##
## Coefficients:
##   SL.glm_All      0
## tmle.SL.dbarts2_All  1
##   SL.glmnet_All    0
##
## Estimation of g (treatment mechanism)
## Procedure: SuperLearner, ensemble
## Model:
##   A ~ SL.glm_All + tmle.SL.dbarts.k.5_All + SL.gam_All
##
## Coefficients:
##   SL.glm_All      0.3813556
## tmle.SL.dbarts.k.5_All  0.6186444
##   SL.gam_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: -0.049489
## Estimated Variance: 0.00042436
## p-value: 0.016288
## 95% Conf Interval: (-0.089865, -0.0091133)
##
## Additive Effect among the Treated
## Parameter Estimate: -0.043794
## Estimated Variance: 0.0003852
## p-value: 0.025656
## 95% Conf Interval: (-0.082262, -0.0053261)
##
## Additive Effect among the Controls
## Parameter Estimate: -0.053913
## Estimated Variance: 0.00046918
## p-value: 0.012812
## 95% Conf Interval: (-0.096367, -0.011458)
##
## Relative Risk
## Parameter Estimate: 0.44923
## p-value: 0.017903
## 95% Conf Interval: (0.23162, 0.8713)
##
## log(RR): -0.80022
## variance(log(RR)): 0.11424
##
## Odds Ratio
## Parameter Estimate: 0.42606
## p-value: 0.016607
## 95% Conf Interval: (0.21197, 0.85637)
##
## log(OR): -0.85317
## variance(log(OR)): 0.12687
```

```
#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.

```
modelo2 <- tmle(Y=bajas_4$READAPTADO,
               A=bajas_4$smmental,
               W=bajas_4[,c(7,8,9,10)],
               id=bajas_4$Num_Unico,
               family="binomial")
summary(modelo2)
```

```
## Initial estimation of Q
## Procedure: cv-SuperLearner, ensemble
## Model:
## Y ~ SL.glm_All + tmle.SL.dbarts2_All + SL.glmnet_All
```

```

##
## Coefficients:
##      SL.glm_All      0.7919848
## tmle.SL.dbarts2_All  0.2080152
##      SL.glmnet_All    0
##
## Estimation of g (treatment mechanism)
## Procedure: SuperLearner, ensemble
## Model:
##      A ~ SL.glm_All + tmle.SL.dbarts.k.5_All + SL.gam_All
##
## Coefficients:
##      SL.glm_All      0.6605538
## tmle.SL.dbarts.k.5_All  0.3394462
##      SL.gam_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: 0.04337
## Estimated Variance: 0.00066595
## p-value: 0.092841
## 95% Conf Interval: (-0.0072102, 0.093949)
##
## Additive Effect among the Treated
## Parameter Estimate: 0.04831
## Estimated Variance: 0.00063243
## p-value: 0.054732
## 95% Conf Interval: (-0.00098074, 0.0976)
##
## Additive Effect among the Controls
## Parameter Estimate: 0.040898
## Estimated Variance: 0.00066701
## p-value: 0.1133
## 95% Conf Interval: (-0.0097225, 0.091518)
##
## Relative Risk
## Parameter Estimate: 1.0915
## p-value: 0.083668
## 95% Conf Interval: (0.98842, 1.2053)
##
## log(RR): 0.087536
## variance(log(RR)): 0.0025609
##
## Odds Ratio
## Parameter Estimate: 1.1896
## p-value: 0.093079
## 95% Conf Interval: (0.97141, 1.4567)

```

```
##
##           log(OR):  0.1736
##      variance(log(OR)):  0.010686
```

```
#print(modelo)
```

## 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$READAPTADO, A=bajas_4$smmental, W=bajas_4[,c(7,8,9,10)], id=bajas_4$Num_Unico,
               family="binomial", g.SL.library = SL.library.2)
```

```
## Loading required namespace: randomForest
```

```
## Loading required namespace: rpart
```

```
summary(modelo3)
```

```
## Initial estimation of Q
## Procedure: cv-SuperLearner, ensemble
## Model:
##      Y ~ SL.glm_All + tmle.SL.dbarts2_All + SL.glmnet_All
##
## Coefficients:
##      SL.glm_All      0
##      tmle.SL.dbarts2_All  0.419164
##      SL.glmnet_All      0.580836
##
## Estimation of g (treatment mechanism)
## Procedure: SuperLearner, ensemble
## 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      0.2754068
##      SL.randomForest_All  0.08280362
##      SL.rpart_All      0.6417896
##
## 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:  0.060167
##      Estimated Variance:  0.00060182
```

```

##           p-value: 0.014182
##    95% Conf Interval: (0.012085, 0.10825)
##
## Additive Effect among the Treated
##    Parameter Estimate: 0.055917
##    Estimated Variance: 0.00063215
##           p-value: 0.02615
##    95% Conf Interval: (0.0066372, 0.1052)
##
## Additive Effect among the Controls
##    Parameter Estimate: 0.062682
##    Estimated Variance: 0.00059443
##           p-value: 0.010141
##    95% Conf Interval: (0.014896, 0.11047)
##
## Relative Risk
##    Parameter Estimate: 1.1272
##           p-value: 0.010902
##    95% Conf Interval: (1.0279, 1.2361)
##
##           log(RR): 0.11975
##    variance(log(RR)): 0.0022127
##
## Odds Ratio
##    Parameter Estimate: 1.2725
##           p-value: 0.014468
##    95% Conf Interval: (1.049, 1.5436)
##
##           log(OR): 0.24097
##    variance(log(OR)): 0.00971

```

```
# print(modelo3)
```