

Untitled

Jan Franco Figueredo Roncal

2024-06-29

```
load("eh22.Rdata")
Seleccion de los datos a utilizar.
Utilizaremos la base eh22p, e22e, para este estudio.
persona <- eh22p
persona$s04b_12 <- as.factor(persona$s04b_12)

personas <- persona %>% dplyr::select(pobre=p0,folio,depto,area,jefe=s01a_05,aestudio,
  yhog,antNeg=s04b_11aa,tipaAct=s04b_12,caeb_op)
personas <- personas %>% filter((jefe==1))
personas$tipaAct <- as.factor(personas$tipaAct)
personas <- personas %>%dplyr::filter((tipaAct == "1" | tipaAct == "3"))
head(personas,5)
```

Se busca incluir a los gastos de equipamiento.

```
bdequi <- eh22e %>% dplyr::select(folio,s08b_4) %>%
  dplyr::group_by(folio)%>% summarise(equi_hogar=sum(s08b_4,na.rm = TRUE))

aux <- left_join(personas, bdequi, by = "folio")
md.pattern(aux)
```

Verificando existencia de datos vacios.

```
bd <- na.omit(aux)
bd<-bd %>% dplyr::select(-folio, -jefe) %>% to_factor()
bd<-bd %>% mutate(pobre=(pobre=="Pobre"))
```

Se crea la base de enrenamiento y prueba.

```
set.seed(605)
aux<-createDataPartition(bd$pobre, p=0.7 , list = F)
bdtrain<-bd[aux,]
bdtest<-bd[-aux,]
```

```
m1<-glm(pobre ~ . , data=bdtrain, family = binomial(link="logit"))
m2<-glm(pobre ~ . , data=bdtrain, family = binomial(link="probit"))
```