

Tarea billetes con restricción

Alejandro Hernández Farías, C.U. 87806

Realizamos el mismo ejercicio del examen pero ahora considerando un parámetro β_j distinto para cada denominación e incluyendo un intercepto. Para lo anterior es necesario incorporar la siguiente restricción en el modelo:

$$\sum_j \beta_j = 0$$

Utilizaremos de nueva cuenta la liga logística y cloglog.

Modelo lineal generalizado binomial con liga logística

Las estimadores puntuales son:

```
##      a.adj      b.adj[1]      b.adj[2]      b.adj[3]      b.adj[4]      b.adj[5]
## -4.30077669 -1.93348682  1.05656813  0.58734450  0.22470406  0.06487014
```

Los estimadores por intervalo son:

```
##              2.5%       97.5%
## a.adj      -4.3571047 -4.2492803
## b.adj[1]   -2.1436303 -1.7507969
## b.adj[2]    0.9918478  1.1234041
## b.adj[3]    0.5192725  0.6575625
## b.adj[4]    0.1561308  0.2955597
## b.adj[5]   -0.0115018  0.1441194
```

Todos los coeficientes con excepcion del correspondiente a los billetes de 500 pesos son significativos.

El DIC es el siguiente

```
## [1] 735781.7
```

La pseudo R^2 de Jags:

```
## [1] 0.5544669
```

Modelo lineal generalizado binomial con liga cloglog

Las estimadores puntuales son:

```
##      a.adj      b.adj[1]      b.adj[2]      b.adj[3]      b.adj[4]      b.adj[5]
## -4.31076859 -1.92469328  1.04682882  0.58428554  0.22534854  0.06823038
```

Los estimadores por intervalo son:

```
##              2.5%       97.5%
## a.adj      -4.365015124 -4.2588838
## b.adj[1]   -2.121746906 -1.7381343
## b.adj[2]    0.982729068  1.1121620
## b.adj[3]    0.516212677  0.6534332
## b.adj[4]    0.156921581  0.2950451
## b.adj[5]   -0.008705263  0.1458145
```

Todos los coeficientes con excepcion del correspondiente a los billetes de 500 pesos son significativos.

El DIC es el siguiente

```
## [1] 1296.73
```

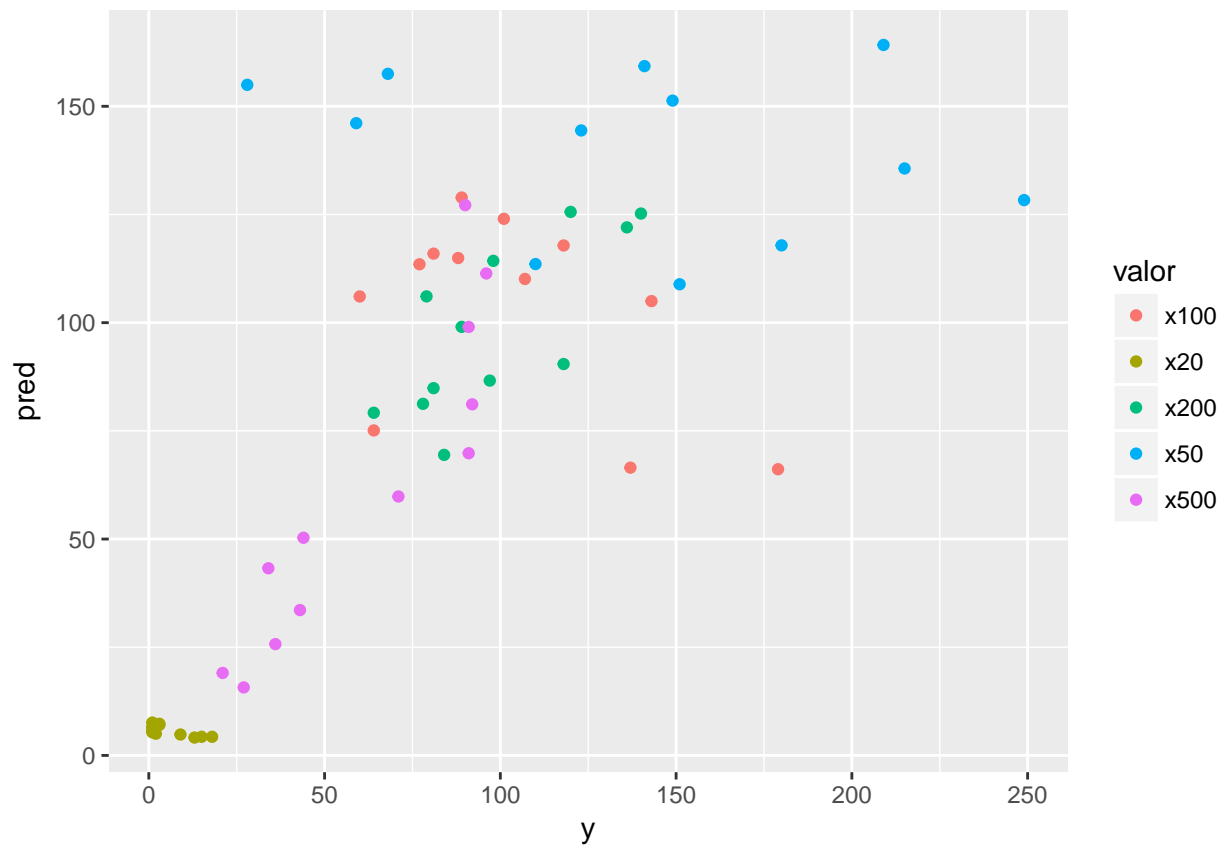
La pseudo R^2 de Jags:

```
## [1] 0.5544175
```

Tabla comparativa de modelos

```
##      Modelo      Link      DIC      R2
## [1,] "Binomial" "logistico" "735782" "0.55447"
## [2,] "Binomial" "cloglog"  "1297"   "0.55442"
```

El mejor modelo es el que tiene funcion liga cloglog, por lo que realizaremos una gráfica para comparar la predicción de numero de billetes falsos utilizando este modelo:



Finalmente, podemos constatar que también existe gran dispersión en este caso, particularmente en los billetes de 50 y 100 pesos.

Anexos

Codigo R

```
library(tidyverse)
library(R2jags)
```

```

prob<-function(x){
  out<-min(length(x[x>0])/length(x),length(x[x<0])/length(x))
  out
}
billetes<-read.csv("./BillsMXc.csv",header=TRUE)

## Liga Logística

n<-dim(billetes)[1]

# Definimos los datos.
y<-billetes$Y
c<-billetes$C
x20<-billetes$x20
x50<-billetes$x50
x100<-billetes$x100
x200<-billetes$x200
x500<-billetes$x500
data<-list("n"=n,"y"=y,"e"=c,"x20"=x20,"x50"=x50,"x100"=x100,"x200"=x200,"x500"=x500)

# Inicializamos los inits
inits<-function(){list(a=1,b=rep(0,5),yf=rep(0,n))}

# Definimos los parámetros que queremos revisar
parameters<-c("a.adj","b.adj","yf")

#Corremos el modelo
set.seed(3452345)
tarea_logit<-jags(data,inits,parameters,model.file="tarea_logit.txt",
  n.iter=300000,n.chains=2,n.burnin=30000,n.thin=1)

#Estimadores puntuales
tarea_logit$BUGSoutput$summary[1:6,1]

#Intervalos
tarea_logit$BUGSoutput$summary[1:6,c(3,7)]

#DIC
tarea_logit$BUGSoutput$DIC

#R2
out.logistico<-tarea_logit$BUGSoutput$summary
out.yf<-out.logistico[grepl("yf",rownames(out.logistico)),]

r2_logistico<-cor(y,out.yf[,1])^2
r2_logistico

## Liga cloglog
# Inicializamos los inits
inits<-function(){list(a=1,b=rep(0,5),yf=rep(0,n))}

```

```

# Definimos los parámetros que queremos revisar
parameters<-c("a.adj","b.adj","yf")

#Corremos el modelo
set.seed(3452345)
tarea_cloglog<-jags(data,inits,parameters,model.file="tarea_cloglog.txt",
                    n.iter=300000,n.chains=2,n.burnin=30000,n.thin=1)

#Estimadores puntuales
tarea_cloglog$BUGSoutput$summary[1:6,1]

#Intervalos
tarea_cloglog$BUGSoutput$summary[1:6,c(3,7)]

#DIC
tarea_cloglog$BUGSoutput$DIC

#R2
out.cloglog<-tarea_cloglog$BUGSoutput$summary
out.yf<-out.cloglog[grepl("yf",rownames(out.cloglog)),]

r2_cloglog<-cor(y,out.yf[,1])^2
r2_cloglog

## Comparativa modelos
Modelo<-c("Binomial","Binomial")
Link<-c("logistico","cloglog")
DIC<-c(round(tarea_logit$BUGSoutput$DIC,0),round(tarea_cloglog$BUGSoutput$DIC,0))
R2<-c(round(r2_logistico,5),round(r2_cloglog,5))
comparativo<-cbind(Modelo,Link,DIC,R2)
comparativo

## Gráfica predicciones
billetes_denom<- gather(billetes,valor,ref,x20:x500)
billetes_pos<-filter(billetes_denom,ref==1)
billetes_final<-mutate(billetes_pos,prop=Y/C)
sum_e<-tarea_cloglog$BUGSoutput$summary
pred<-sum_e[grepl("yf",rownames(sum_e)),1]
billetes_final2<-cbind(billetes_final,pred)
ggplot(billetes_final2)+geom_point(aes(y,pred,color=valor))

```

Codigo Jags

```

## "tarea_logit.txt"
model
{
  #Likelihood
  for (i in 1:n){
    y[i] ~ dbin(pi[i],e[i])
    logit(pi[i])<-a+b[1]*x20[i]+b[2]*x50[i]+b[3]*x100[i]+b[4]*x200[i]+b[5]*x500[i]
  }
}

```

```

#Priors
a ~ dnorm(0,0.001)
for (j in 1:5){
  b[j] ~ dnorm(0,0.001)
}

#Estimable parameters
a.adj<-a+mean(b[])
for (j in 1:5){b.adj[j]<-b[j]-mean(b[])}

#Predicción
for (i in 1:n){
  yf[i] ~ dbin(pif[i],e[i])
  logit(pif[i])<-a.adj+b.adj[1]*x20[i]+b.adj[2]*x50[i]+b.adj[3]*x100[i]+b.adj[4]*x200[i]+b.adj[5]*x500[i]
}

}

## "tarea_cloglog.txt"
model
{
#Likelihood
for (i in 1:n){
  y[i] ~ dbin(pi[i],e[i])
  cloglog(pi[i])<-a+b[1]*x20[i]+b[2]*x50[i]+b[3]*x100[i]+b[4]*x200[i]+b[5]*x500[i]
}

#Priors
a ~ dnorm(0,0.001)
for (j in 1:5){
  b[j] ~ dnorm(0,0.001)
}

#Estimable parameters
a.adj<-a+mean(b[])
for (j in 1:5){b.adj[j]<-b[j]-mean(b[])}

#Predicción
for (i in 1:n){
  yf[i] ~ dbin(pif[i],e[i])
  cloglog(pif[i])<-a.adj+b.adj[1]*x20[i]+b.adj[2]*x50[i]+b.adj[3]*x100[i]+b.adj[4]*x200[i]+b.adj[5]*x500[i]
}

}

```