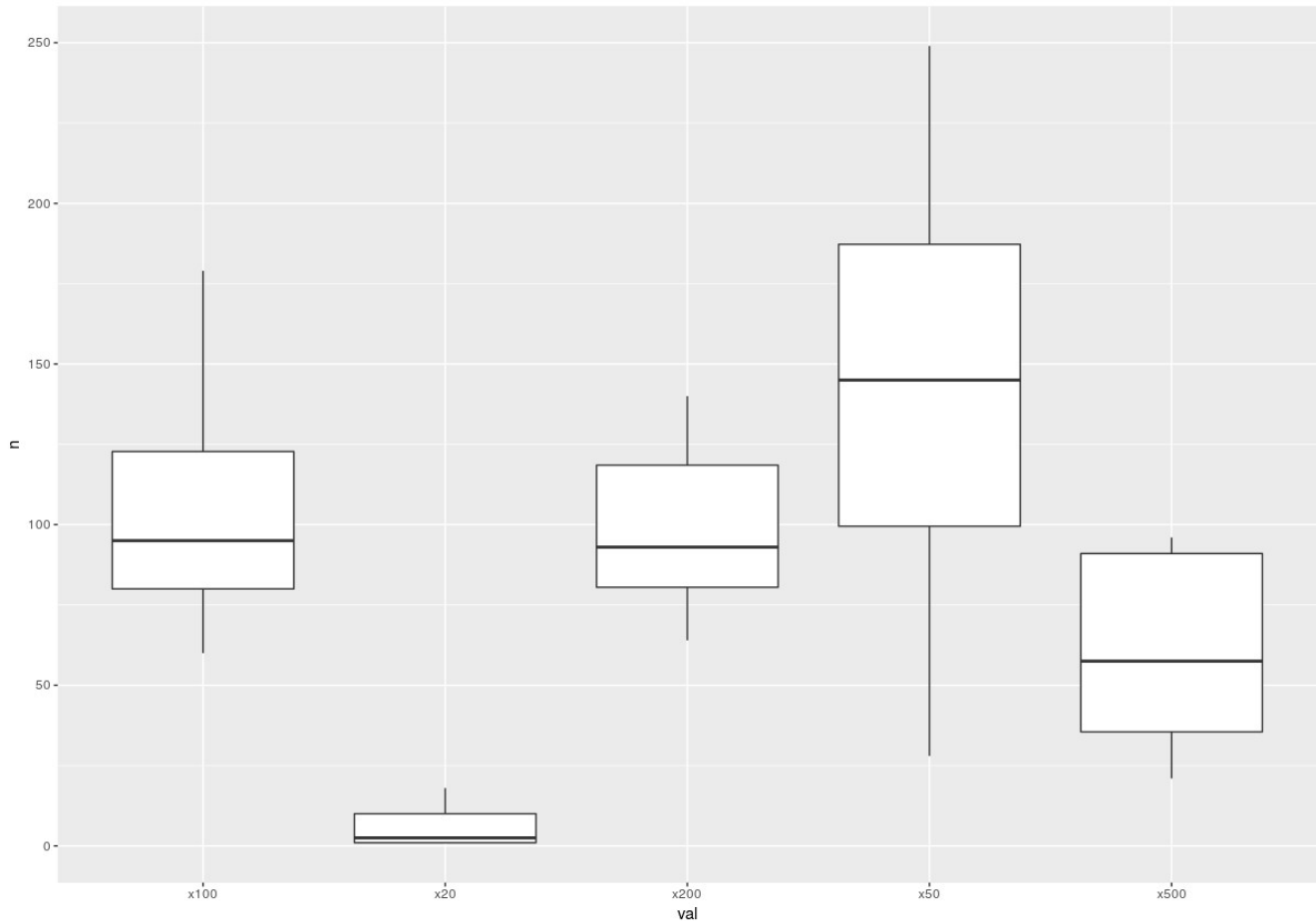


1. El Banco de México es el responsable de emitir los billetes que circulan en la economía mexicana. Se cuenta con la información del número de billetes en circulación (C) y la cantidad de billetes falsos (Y), ambas en millones de piezas, para los años de 2000 a 2011. Para identificar la denominación del billete definimos variables indicadoras x20, x50, x100, x200 y x500. La base de datos se encuentra en [http://allman.rhon.itam.mx/~lnieto/index\\_archivos/BillsMXc.csv](http://allman.rhon.itam.mx/~lnieto/index_archivos/BillsMXc.csv)

a) Realiza un análisis exploratorio de los datos. Crea las gráficas y encuentra las estadísticas que mejor describan la información y coméntalas. Obtén conclusiones por tipo de denominación.



Se hizo una gráfica de caja y brazos donde se muestra en el eje de las xs la denominación de los billetes vs. el número de billetes falsos. Cada caja representa la distribución del número de billetes falsos por año. Podemos observar que el billete más falsificado es el de 50 ya que podemos ver claramente que su mediana es de casi 150 millones de billetes falsos, y al menos 75% de los años se han falsificado más de 100 millones de billetes de 50, el que menos se falsifica es el billete de 20, el segundo menos falsificado es el de 500 con mediana arriba de 50 millones de piezas pero ningún año pasó los 100 millones de piezas, el de 200 y el de 100 son parecidos en sus medianas cerca de 100 millones.

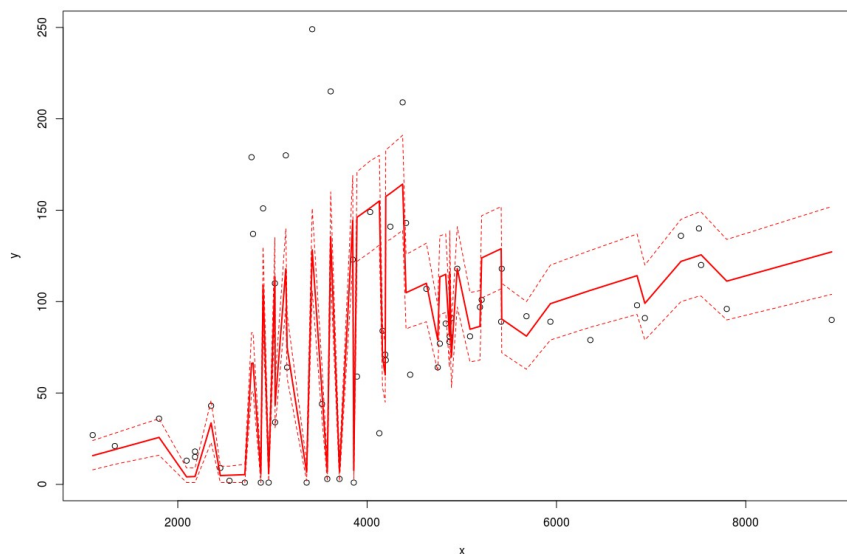
b) Ajuste el modelo de regresión binomial con liga logística, i.e., como distribuciones iniciales  $\sim 0,0.001$  para 1, ... ,5. Calcula los indicadores de ajuste DIC y pseudoR2. Encuentra los estimadores puntuales y por intervalo de los parámetros del modelo, interprétalos y comenta qué tan bueno es el modelo.

DIC=[1] 1297  
R2=1] 0.5542991

	mean	sd	2.5%	25%	50%	75%	97.5%
Rhat n.eff							
b[1]	-6.23353189	1.218929e-01	-6.479000	-6.313000	-6.231000	-6.151e+00	-6.002e+00
1.000963 18000							
b[2]	2.98884361	1.243028e-01	2.754000	2.904000	2.986000	3.071e+00	3.239e+00
1.000961 18000							
b[3]	2.51902294	1.249656e-01	2.279975	2.435000	2.517000	2.601e+00	2.771e+00
1.000987 18000							
b[4]	2.15625694	1.247832e-01	1.918975	2.072000	2.153500	2.238e+00	2.405e+00
1.000948 18000							
b[5]	1.99839500	1.274431e-01	1.754975	1.912000	1.996000	2.082e+00	2.252e+00
1.000953 18000							

los valores de las betas se ueden ver en amarillo donde los intervalos de cfza an del percentil 2.5% al 97.5%. y la esimación es la media.

podemos ver que el ajuste del modelo no es muy bueno dao que la r2 es muy baja entonces el modelo explica poca variabilidad de los datos además en la gráfica se ve que muchas predicciones están fuera del intervalo de confianza.



c) En el modelo de regresión binomial con liga logística, ¿cuál es la interpretación del coeficiente  $b_1$  en el modelo?, ¿cómo interpretas la suma  $b_1 + b_j$  para  $j = 2, \dots, 5$ ?

$b_1$  representa el valor del coeficiente para la variable de billetes de 20 que es nuestra base dado que no está en el modelo explícita, ....

Anexo código bugs:

```
model{
#Likelihood
for (i in 1:n){
y[i]~dbin(pi[i],e[i])

logit(pi[i])<-b[1]+b[2]*x50[i]+b[3]*x100[i]+b[4]*x200[i]+b[5]*x500[i]

}

#Priors

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

for (i in 1:n) { yf[i] ~ dbin(pi[i],e[i])}

}
```

Anexo código R:

```
library(R2OpenBUGS)
library(R2jags)
```

```
wdir<-"/home/abraham/RA2018/ejemex/ex2018"
```

```
setwd(wdir)
```

```
#--- Funciones utiles ---
```

```
prob<-function(x){  
  out<-min(length(x[x>0])/length(x),length(x[x<0])/length(x))  
  out  
}
```

```
#Leemos datos
```

```
bill<-
```

```
read.table("http://allman.rhon.itam.mx/~lnieto/index_archivos/BillsMXc.csv",sep=";",header=TRUE)
```

```
bill
```

```
#x<-bill$C#/1000
```

```
y<-bill$Y#/1000
```

```
library(tidyverse)
```

```
library(ggplot2)
```

```
datan<-gather(bill,val,I,x20:x500) %>%filter(.,I==1) %>%mutate(.,proporcion_falsos=Y/C)
```

```
explo <- ggplot(datan,aes(x=val,y=Y))+geom_boxplot()+xlab("val")+ylab("n")
```

```
plot(x,y,type = "p")
```

```
plot(bill$Year,y)
```

```
#hist(bill$Year)
```

```
cor(x,y)
```

```
n<-nrow(bill)
```

```
#-Defining data-
```

```
#data<-list(x1=sal$X1,x2=sal$X2,x3=sal$X3)
```

```
data<-
```

```
list("n"=n,"y"=y,"e"=bill$C,"x50"=bill$x50,"x100"=bill$x100,"x200"=bill$x200,"x500"=bill$x500)
```

```
#-Defining inits-
```

```
inits<-function(){list(b=rep(0,5),yf=rep(0,n))}
```

```
parameters<-c("b","yf","pi")
```

```
#OpenBUGS
```

```
ex1<-bugs(data,inits,parameters,model.file="eja.txt",  
          n.iter=10000,n.chains=2,n.burnin=1000)
```

```
#ex1j<-jags(data,inits,parameters,model.file="eja.txt",
```

```
#          n.iter=50000,n.chains=1,n.burnin=5000,n.thin = 1)
```

```

res<-function(func,out){
  ex1.sim<-func
  #out<-ex1.sim$sims.list

  for(i in 1:5){
    b<-out$b[,i]
    par(mfrow=c(1,1))
    plot(b,type="l",xlab = "beta")
    plot(cumsum(b)/(1:length(b)),type="l",xlab = "beta")
    hist(b,freq=FALSE, main = "beta")
    acf(b)
  }
  plot(out$b)

  #out.sum<-ex1.sim$summary
  #return(ex1.sim)

}

rex<-function(fun){return(fun)}

out<-rex(ex1)$sims.list

out.sum<-rex(ex1)$summary

res(ex1,out)

#Predictions
adj<-function(out.sum,y,a){
  out.yf<-out.sum[grepl(a,rownames(out.sum)),]
  or<-order(x)
  ymin<-min(y,out.yf[,c(1,3,7)])
  ymax<-max(y,out.yf[,c(1,3,7)])
  par(mfrow=c(1,1))
  plot(x,y,ylim=c(ymin,ymax))
  lines(x[or],out.yf[or,1],lwd=2,col=2)
  lines(x[or],out.yf[or,3],lty=2,col=2)
  lines(x[or],out.yf[or,7],lty=2,col=2)

  plot(out.yf[1:n,1],y,type="p")

}

adj(out.sum,y,"yf")

```

```
#DIC
met<-function(ex1.sim,out.sum){
  out.dic<-ex1.sim$DIC
  #out.dic<-ej4.sim$BUGSoutput$DIC
  print(out.dic)
  print(out.sum)
}

met(ex1,out.sum)

r2<-function(y,out.sum,a){
  out.yf<-out.sum[grepl(a,rownames(out.sum)),]
  R2<-(cor(y,out.yf[1:n,1]))^2
  print(R2)
}

r2(y,out.sum,"yf")
```