

Examen Parcial

Eduardo Hidalgo, eduardohidalgogarcia@gmail.com

October 27, 2018

Definición del problema

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

```
knitr::kable(head(bills), caption = "Base de datos")
```

Table 1: Base de datos

Year	C	Y	x20	x50	x100	x200	x500
2000	2182	15	1	0	0	0	0
2001	2093	13	1	0	0	0	0
2002	2182	18	1	0	0	0	0
2003	2449	9	1	0	0	0	0
2004	2546	2	1	0	0	0	0
2005	2708	1	1	0	0	0	0

Inciso (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.

La base se encuentra estructurada de forma vertical, bloques de 12 en 12 observaciones:

```
knitr::kable(bills%>%group_by(Year)%>%summarise(CT=sum(C),YT=sum(Y),Veces20=sum(x20),Veces50=sum(x50),Veces100=sum(x100),Veces200=sum(x200),Veces500=sum(x500)))
```

Table 2: Estructura de los datos

Year	CT	YT	Veces20	Veces50	Veces100	Veces200	Veces500
2000	13366	485	1	1	1	1	1
2001	13870	386	1	1	1	1	1
2002	15357	325	1	1	1	1	1
2003	18371	331	1	1	1	1	1
2004	19645	384	1	1	1	1	1
2005	19484	555	1	1	1	1	1
2006	21245	483	1	1	1	1	1
2007	22824	371	1	1	1	1	1
2008	24618	338	1	1	1	1	1
2009	26829	339	1	1	1	1	1
2010	28419	408	1	1	1	1	1
2011	30085	509	1	1	1	1	1

Cada bloque esta basado en la denominación indicada por la variable del tipo `x` cuyo valor sea igual a 1. En él se consolida la información del año (`Year`), la cantidad de billetes por denominacion en circulación (`C`) y la cantidad de éstos falsos (`Y`).

Para entender la base es necesario llevarla de un formato *long* a uno *wide*. Así podremos entender el resumen de las variables, a traves de la tasa o proporción de billetes falsos por tipo de denonminación, y ver que en realidad contamos con 5 variables cada una con 12 observaciones

```
x20dba<-bills%>%filter(x20==1)%>%select(Year,C,Y)%>%mutate(PropFalsos=Y/C)
colnames(x20dba)<-c("Year","Cantidad_20","Falsos_20","PropFalsos_20")
x50dba<-bills%>%filter(x50==1)%>%select(Year,C,Y)%>%mutate(PropFalsos=Y/C)
colnames(x50dba)<-c("Year","Cantidad_50","Falsos_50","PropFalsos_50")
x100dba<-bills%>%filter(x100==1)%>%select(Year,C,Y)%>%mutate(PropFalsos=Y/C)
colnames(x100dba)<-c("Year","Cantidad_100","Falsos_100","PropFalsos_100")
x200dba<-bills%>%filter(x200==1)%>%select(Year,C,Y)%>%mutate(PropFalsos=Y/C)
colnames(x200dba)<-c("Year","Cantidad_200","Falsos_200","PropFalsos_200")
x500dba<-bills%>%filter(x500==1)%>%select(Year,C,Y)%>%mutate(PropFalsos=Y/C)
colnames(x500dba)<-c("Year","Cantidad_500","Falsos_500","PropFalsos_500")
billswide<-inner_join(x20dba,x50dba,by="Year")%>%inner_join(x100dba,by="Year")%>%inner_join(x200dba,by="Year")%>%inner_join(x500dba,by="Year")
billswide_final<-billswide%>%select(Year,PropFalsos_20,PropFalsos_50,PropFalsos_100,PropFalsos_200,PropFalsos_500)
knitr::kable(summary(billswide_final), caption = "Resuemn de Proporción de Billetes Falsos x Denominación")
```

Table 3: Resuemn de Proporción de Billetes Falsos x Denominación

Year	PropFalsos_20	PropFalsos_50	PropFalsos_100	PropFalsos_200	PropFalsos_500
Min. :2000	Min. :0.0002591	Min. :0.006781	Min. :0.01347	Min. :0.01243	Min. :0.01010
1st Qu.:2003	1st Qu.:0.0003451	1st Qu.:0.028024	1st Qu.:0.01658	1st Qu.:0.01482	1st Qu.:0.01244
Median :2006	Median :0.0007973	Median :0.036653	Median :0.01984	Median :0.01598	Median :0.01596
Mean :2006	Mean :0.0024212	Mean :0.038836	Mean :0.02611	Mean :0.01675	Mean :0.01579
3rd Qu.:2008	3rd Qu.:0.0043090	3rd Qu.:0.053365	3rd Qu.:0.02598	3rd Qu.:0.01866	3rd Qu.:0.01836
Max. :2011	Max. :0.0082493	Max. :0.072807	Max. :0.06441	Max. :0.02176	Max. :0.02452

```
#knitr::kable(summary(inner_join(x20dba,x50dba,by="Year")), caption = "Resumen de Variables Baja Denominación")
```

```
#knitr::kable(summary(inner_join(x100dba,x200dba,by="Year")%>%inner_join(x500dba,by="Year")), caption = "Resumen de Variables Alta Denominación")
```

Y la proporción de billetes falsos del total de los billetes en circulación :

```
knitr::kable(summary(billswide%>%select(PropFalsos_Total)), caption = "Resumen de Proporción Total de Billetes Falsos")
```

Table 4: Resumen de Proporción Total de Billetes Falsos

PropFalsos_Total
Min. :0.01264
1st Qu.:0.01578
Median :0.01878
Mean :0.02066
3rd Qu.:0.02401
Max. :0.03629

Podemos observar que en promedio los billetes de 50 mxn son los que tienen mayor proporción de billetes falsos. Por arriba del promedio general de la proporción de billetes falsificados tenemos a los billetes de 20,

50 y 100 mxn. Por debajo del promedio general tenemos a los billetes de 200 y 500 mxn.

El modelo

Ignorando la dependencia temporal que pudiera existir entre las observaciones de distintos años, considera un modelo de regresión binomial de la forma

$$Y \sim \text{Bin}(C_i, \pi_i) \text{ donde } i = 1, \dots, 60$$

define el predictor lineal: $\eta_i = \beta_1 + \beta_2 x_{50i} + \beta_3 x_{100i} + \beta_4 x_{200i} + \beta_5 x_{500i}$

Nota la variable x_{20i} no esta presente en el predictor lineal pues no queremos se indetermine el modelo.

EL objetivo es **comparar** el desempeño del **modelo lineal generalizado binomial** bajo **dos funciones** **liga**: la **logística** y la complementaria **log-log**.

Inciso (b)

Ajuste el modelo de regresión binomial con liga logística, i.e., $\text{logit}(\pi_i) = \eta_i$ y usa como distribuciones iniciales $\beta_j \sim N(0, 0.001)$ donde $j = 1, \dots, 5$. 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.

```
bills_model<-bills%>%mutate(Prop_Falsos=Y/C)

n<-nrow(bills_model)

data<-list("n"=n,"ne"=bills_model$C,"y"=bills_model$Y,"x1"=bills_model$x50,"x2"=bills_model$x100,"x3"=bills_model$x200,"x4"=bills_model$x500)

#-Defining inits-
#inits<-function(){list(alpha=1,beta=1,tau=1,yf=rep(0,n))}
inits<-function(){list(beta=rep(0,5),yf=rep(0,n))}

#-Selecting parameters to monitor-
#parameters<-c("alpha","beta","tau","yf")
parameters<-c("beta","p","yf")

#-Running code-
#OpenBUGS
ejb.sim.bugs<-bugs(data,inits,parameters,model.file="Ejb.txt",
                  n.iter=10000,n.chains=1,n.burnin=1000)

#JAGS
ejb.sim.jags<-jags(data,inits,parameters,model.file="Ejb.txt",
                  n.iter=10000,n.chains=1,n.burnin=1000,n.thin=1)

#ejb.sim.jags<-jags(dataH,inits,parameters,model.file="Ejb1.txt",
#                  n.iter=10000,n.chains=1,n.burnin=1000,n.thin=1)
#OpenBUGS
out.bugs<-ejb.sim.bugs$sims.list

#JAGS
out.jags<-ejb.sim.jags$BUGSoutput$sims.list
```

```

#Resumen (estimadores)
#OpenBUGS
out.sum.bugs<-ejb.sim.bugs$summary

#JAGS
out.sum.jags<-ejb.sim.jags$BUGSoutput$summary

#Tabla resumen BUGS
out.sum.t<-out.sum.bugs[grepl("beta",rownames(out.sum.bugs)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out.bugs$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
out.sum.t.bugs<-out.sum.t
knitr::kable(out.sum.t.bugs, caption = "Estimadores Puntuales e Intervalos de Confianza: BUGS",row.names=)

```

Table 5: Estimadores Puntuales e Intervalos de Confianza: BUGS

	mean	2.5%	97.5%	prob
beta[1]	-5.870825	-6.210025	-5.090	0
beta[2]	2.620052	1.871000	2.979	0
beta[3]	2.153162	1.373000	2.515	0
beta[4]	1.791200	1.046000	2.146	0
beta[5]	1.632726	0.884700	1.999	0

```

#Tabla resumen JAGS
out.sum.t<-out.sum.jags[grepl("beta",rownames(out.sum.jags)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out.jags$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
out.sum.t.jags<-out.sum.t
knitr::kable(out.sum.t.jags, caption = "Estimadores Puntuales e Intervalos de Confianza: JAGS",row.names=)

```

Table 6: Estimadores Puntuales e Intervalos de Confianza: JAGS

	mean	2.5%	97.5%	prob
beta[1]	-6.212690	-6.470050	-6.014519	0
beta[2]	2.967543	2.765230	3.226084	0
beta[3]	2.497611	2.292354	2.757480	0
beta[4]	2.135308	1.928477	2.398113	0
beta[5]	1.977390	1.770913	2.241088	0

```

ModeloLigaLogit<-out.sum.t.jags
CompletoLigaLogit<-out.sum.jags
#DIC
out.dic.bugs<-ejb.sim.bugs$DIC
out.dic.jags<-ejb.sim.jags$BUGSoutput$DIC
#Pseudo R
aux<-out.sum.bugs[grepl("yf",rownames(out.sum.bugs)),1]%>%as.data.frame()
colnames(aux)[1]<-c("ygorro")
R2.bugs<-(cor(bills_model$Y,aux$ygorro))^2

aux<-out.sum.jags[grepl("yf",rownames(out.sum.jags)),1]%>%as.data.frame()
colnames(aux)[1]<-c("ygorro")

```

```
R2.jags<-(cor(bills_model$Y,aux$ygorro))^2
DIC<-c(out.dic.bugs,out.dic.jags)
PseudoR<-c(R2.bugs,R2.jags)
Software<-c("BUGS","JAGS")
BondadAjusteLigaLogit<-data.frame(Software,PseudoR,DIC)
knitr::kable(BondadAjusteLigaLogit, caption = "Resumen de Medidas de Bondad de Ajuste: BUGS vs. JAGS",r
```

Table 7: Resumen de Medidas de Bondad de Ajuste: BUGS vs. JAGS

Software	PseudoR	DIC
BUGS	0.5507638	1332.000
JAGS	0.5545932	1296.415

En conclusión, la correlación al cuadrado entre lo observado mediante la variable Y y las predicciones hechas con el modelo lineal generalizado binomial con liga logit \hat{Y} fue de .55 un valor muy alejado de una correlación perfecta . Lo anterior, indica que es necesario seguir trabajando con el modelo para mejorar la precisión de nuestras predicciones.

Inciso (c)

c.1

En el modelo de regresión binomial con liga logística, ¿cuál es la interpretación del coeficiente β_1 en el modelo?

En el modelo de regresión binomial con liga logística los coeficientes $\beta_i \forall i \in [1, 5]$ implican la determinación de la probabilidad de éxito en la distribución binomial de la siguiente manera:

$$\begin{aligned}\eta_i &= \beta_1 + \beta_2 x_{50i} + \beta_3 x_{100i} + \beta_4 x_{200i} + \beta_5 x_{500i} \\ \Rightarrow p_i &= \frac{\exp(\eta_i)}{(1 + \exp(\eta_i))} \\ \Rightarrow Y_i &\sim \text{Bin}(p_i, C_i)\end{aligned}$$

En este sentido nota que si $x_{50i} = x_{100i} = x_{200i} = x_{500i} = 0$ entonces se tiene $n_i = \beta_1$, por lo que:

$$\hat{Y}_i = E(Y_i | p_i, c_i, X_i) = E(Y_i | p_i, c_i, x_{20i}) = \beta_{1,i} C_i \text{ donde } i \in [1, 12]$$

Y , \hat{Y}_i denota la cantidad esperada (promedio) de billetes de 20 pesos falsos como resultado de una cantidad total de billetes de 20 pesos emitidos. En realidad el estimador como está planteado sólo tiene sentido para el rango de valores donde $i \in [1, 12]$ pues para los restantes i – valores la variable C_i está representando la cantidad de billetes emitidos de otras denominaciones (en ese caso: $\hat{Y}_i = (\beta_1 + \beta_j) C_i$).

En el caso de la denominación de 20 pesos, se tiene (JAGS):

$$\beta_1 = -6.2$$

$$\Rightarrow \eta_i = \beta_1 = -6.2, \forall i \in [1, 12] \equiv \eta_{20}$$

$$\Rightarrow p_{20} = \frac{\exp(-6.2)}{1 + \exp(-6.2)}$$

```
beta1<-out.sum.jags[grep("beta",rownames(out.sum.jags)),c(1)][1]
p20<-c(exp(beta1)/(1+exp(beta1)))%>%as.data.frame()
colnames(p20)[1]<-c("Prob de Billetes Falsos")
```

```
knitr::kable(p20, caption = "Estimación de la probabilidad de billetes de 20 mxn falsos liga: logit",row
```

Table 8: Estimación de la probabilidad de billetes de 20 mxn falsos
liga: logit

Prob de Billetes Falsos
0.0019998

c.2

En el modelo de regresión binomial con liga logística, ¿cómo interpretas la suma $\beta_i + \beta_j$ para $j = 2, \dots, 5$?

La respuesta es que corresponde al **cambio del momio asociado a la probabilidad de tener billetes falsos, al incluir la denominación j**. Se demuestra de la siguiente manera:

Sea

$$\log\left(\frac{p_i}{1-p_i}\right) = \begin{cases} \beta_1 & \text{si } i \in [1, 12] \\ \beta_1 + \beta_2 & \text{si } i \in [13, 24] \\ \beta_1 + \beta_3 & \text{si } i \in [25, 36] \\ \beta_1 + \beta_4 & \text{si } i \in [37, 48] \\ \beta_1 + \beta_5 & \text{si } i \in [49, 60] \end{cases}$$

donde:

$\frac{p_i}{1-p_i}$ es el momio asociado a la probabilidad de billetes falsos en cada denominacion

Definamos:

$$i = \{20\} \text{ y } j \in \{\{50\}, \{100\}, \{200\}, \{500\}\}$$

$$\Rightarrow \log\left(\frac{p_i}{1-p_i}\right) = \beta_i \equiv \beta_{20} \text{ (caso base)}$$

$$\Rightarrow \log\left(\frac{p_j}{1-p_j}\right) = \beta_1 + \beta_d \text{ donde: } d \in \{\{50\}, \{100\}, \{200\}, \{500\}\}$$

$$\Rightarrow \beta_d = \log\left(\frac{p_j}{1-p_j}\right) - \log\left(\frac{p_i}{1-p_i}\right)$$

$$\Rightarrow \exp(\beta_d) = \frac{\frac{p_j}{1-p_j}}{\frac{p_i}{1-p_i}} \equiv \frac{\text{momio asociado a la probabilidad de billetes falsos distintos a 20 mxn}}{\text{momio asociado a la probabilidad de billetes falsos de 20 mxn}} \equiv \Delta_{\text{momios}}$$

Inciso (d)

En el modelo de regresión binomial con liga logística define “la tasa de billetes falsos por mil circulando” para cada denominación como:

$$p_j = 1000 \frac{\exp(\beta_1 + \beta_j)}{1 + \exp(\beta_1 + \beta_j)}$$

Estima estas tasas mediante un intervalo de 95% de probabilidad y coméntalas.

```
EstimacionP1000_logit<-out.sum.jags[grep("p",rownames(out.sum.jags)),c(1,3,7)]%>%as.data.frame()
EstimacionP1000_logit$mean<-EstimacionP1000_logit$mean*1000
EstimacionP1000_logit$`2.5%`<-EstimacionP1000_logit$`2.5%`*1000
EstimacionP1000_logit$`97.5%`<-EstimacionP1000_logit$`97.5%`*1000
knitr::kable(EstimacionP1000_logit, caption = "Estimación de la tasa de billetes falsos por mil circulando")
```

Table 9: Estimación de la tasa de billetes falsos por mil circulando
liga: logit

	mean	2.5%	97.5%
p[1]	2.011924	1.546752	2.437069
p[2]	2.011924	1.546752	2.437069
p[3]	2.011924	1.546752	2.437069
p[4]	2.011924	1.546752	2.437069
p[5]	2.011924	1.546752	2.437069
p[6]	2.011924	1.546752	2.437069
p[7]	2.011924	1.546752	2.437069
p[8]	2.011924	1.546752	2.437069
p[9]	2.011924	1.546752	2.437069
p[10]	2.011924	1.546752	2.437069
p[11]	2.011924	1.546752	2.437069
p[12]	2.011924	1.546752	2.437069
p[13]	37.512022	35.806281	39.356996
p[14]	37.512022	35.806281	39.356996
p[15]	37.512022	35.806281	39.356996
p[16]	37.512022	35.806281	39.356996
p[17]	37.512022	35.806281	39.356996
p[18]	37.512022	35.806281	39.356996
p[19]	37.512022	35.806281	39.356996
p[20]	37.512022	35.806281	39.356996
p[21]	37.512022	35.806281	39.356996
p[22]	37.512022	35.806281	39.356996
p[23]	37.512022	35.806281	39.356996
p[24]	37.512022	35.806281	39.356996
p[25]	23.783714	22.500079	25.113533
p[26]	23.783714	22.500079	25.113533
p[27]	23.783714	22.500079	25.113533
p[28]	23.783714	22.500079	25.113533
p[29]	23.783714	22.500079	25.113533
p[30]	23.783714	22.500079	25.113533
p[31]	23.783714	22.500079	25.113533
p[32]	23.783714	22.500079	25.113533
p[33]	23.783714	22.500079	25.113533
p[34]	23.783714	22.500079	25.113533

	mean	2.5%	97.5%
p[35]	23.783714	22.500079	25.113533
p[36]	23.783714	22.500079	25.113533
p[37]	16.676160	15.734583	17.643933
p[38]	16.676160	15.734583	17.643933
p[39]	16.676160	15.734583	17.643933
p[40]	16.676160	15.734583	17.643933
p[41]	16.676160	15.734583	17.643933
p[42]	16.676160	15.734583	17.643933
p[43]	16.676160	15.734583	17.643933
p[44]	16.676160	15.734583	17.643933
p[45]	16.676160	15.734583	17.643933
p[46]	16.676160	15.734583	17.643933
p[47]	16.676160	15.734583	17.643933
p[48]	16.676160	15.734583	17.643933
p[49]	14.278264	13.266720	15.326897
p[50]	14.278264	13.266720	15.326897
p[51]	14.278264	13.266720	15.326897
p[52]	14.278264	13.266720	15.326897
p[53]	14.278264	13.266720	15.326897
p[54]	14.278264	13.266720	15.326897
p[55]	14.278264	13.266720	15.326897
p[56]	14.278264	13.266720	15.326897
p[57]	14.278264	13.266720	15.326897
p[58]	14.278264	13.266720	15.326897
p[59]	14.278264	13.266720	15.326897
p[60]	14.278264	13.266720	15.326897

Recordemos que la base esta estructurada con base en bloques de 12 en 12 observaciones. Es por esto que podemos observar cambios estructurales en la probabilidad asociada a billetes falsos de 12 en 12 observaciones.

Inciso (e)

Ajuste el modelo de regresión binomial con liga complementaria log-log, i.e., $\log(-\log(1 - p_i)) = \eta_i$ y usa como distribuciones iniciales $\beta_j \sim N(0, 0.001)$ donde $j = 1, \dots, 5$. 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.

```
#bills_model<-bills%>%mutate(Prop_Falsos=Y/C)

n<-nrow(bills_model)

data<-list("n"=n,"ne"=bills_model$C,"y"=bills_model$Y,"x1"=bills_model$x50,"x2"=bills_model$x100,"x3"=bills_model$x150)

#-Defining inits-
#inits<-function(){list(alpha=1,beta=1,tau=1,yf=rep(0,n))}
inits<-function(){list(beta=rep(0,5),yf=rep(0,n))}

#-Selecting parameters to monitor-
#parameters<-c("alpha","beta","tau","yf")
parameters<-c("beta","p","yf")
```



```

#-Running code-
#OpenBUGS
eje.sim.bugs<-bugs(data,inits,parameters,model.file="Eje.txt",
                  n.iter=10000,n.chains=1,n.burnin=1000)

#JAGS
eje.sim.jags<-jags(data,inits,parameters,model.file="Eje.txt",
                  n.iter=10000,n.chains=1,n.burnin=1000,n.thin=1)

#ejb.sim.jags<-jags(dataH,inits,parameters,model.file="Ejb1.txt",
#                  n.iter=10000,n.chains=1,n.burnin=1000,n.thin=1)
#OpenBUGS
out.bugs<-eje.sim.bugs$sims.list

#JAGS
out.jags<-eje.sim.jags$BUGSoutput$sims.list

#Resumen (estimadores)
#OpenBUGS
out.sum.bugs<-eje.sim.bugs$summary

#JAGS
out.sum.jags<-eje.sim.jags$BUGSoutput$summary

#Tabla resumen BUGS
out.sum.t<-out.sum.bugs[grep("beta",rownames(out.sum.bugs)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out.bugs$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
out.sum.t.bugs<-out.sum.t
knitr::kable(out.sum.t.bugs, caption = "Estimadores Puntuales e Intervalos de Confianza: BUGS",row.names=)

```

Table 10: Estimadores Puntuales e Intervalos de Confianza: BUGS

	mean	2.5%	97.5%	prob
beta[1]	-5.412251	-6.4270	-4.27400	0.0000000
beta[2]	2.104306	0.7546	3.17100	0.0118889
beta[3]	1.633316	0.4582	2.70500	0.0000000
beta[4]	1.319297	0.1133	2.34900	0.0000000
beta[5]	1.108691	-0.1745	2.21105	0.3322222

```

#Tabla resumen JAGS
out.sum.t<-out.sum.jags[grep("beta",rownames(out.sum.jags)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out.jags$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
out.sum.t.jags<-out.sum.t
knitr::kable(out.sum.t.jags, caption = "Estimadores Puntuales e Intervalos de Confianza: JAGS",row.names=)

```

Table 11: Estimadores Puntuales e Intervalos de Confianza: JAGS

	mean	2.5%	97.5%	prob
beta[1]	-6.223056	-6.500179	-6.001861	0
beta[2]	2.959122	2.736909	3.241716	0
beta[3]	2.496302	2.268842	2.780086	0
beta[4]	2.137838	1.910998	2.428140	0

	mean	2.5%	97.5%	prob
beta[5]	1.979792	1.746519	2.267366	0

```

ModeloLigacloglog<-out.sum.t.jags
CompletoLigacloglog<-out.sum.jags
#DIC
out.dic.bugs<-ejb.sim.bugs$DIC
out.dic.jags<-ejb.sim.jags$BUGSoutput$DIC
#Pseudo R
aux<-out.sum.bugs[grepl("yf",rownames(out.sum.bugs)),1]%>%as.data.frame()
colnames(aux)[1]<-c("ygorro")
R2.bugs<-(cor(bills_model$Y,aux$ygorro))^2

aux<-out.sum.jags[grepl("yf",rownames(out.sum.jags)),1]%>%as.data.frame()
colnames(aux)[1]<-c("ygorro")
R2.jags<-(cor(bills_model$Y,aux$ygorro))^2
DIC<-c(out.dic.bugs,out.dic.jags)
PseudoR<-c(R2.bugs,R2.jags)
Software<-c("BUGS","JAGS")
BondadAjusteLigacloglog<-data.frame(Software,PseudoR,DIC)
knitr::kable(BondadAjusteLigacloglog, caption = "Resumen de Medidas de Bondad de Ajuste: BUGS vs. JAGS")

```

Table 12: Resumen de Medidas de Bondad de Ajuste: BUGS vs. JAGS

Software	PseudoR	DIC
BUGS	0.5345095	1332.000
JAGS	0.5547261	1296.415

En conclusión, la correlación al cuadrado entre lo observado mediante la variable Y y las predicciones hechas con el modelo lineal generalizado binomial con liga log-log \hat{Y} fue de .55 un valor muy alejado de una correlación perfecta. Lo anterior, indica que es necesario seguir trabajando con el modelo para mejorar la precisión de nuestras predicciones.

Inciso (f)

f.1

En el modelo de regresión binomial con liga log-log, ¿cuál es la interpretación del coeficiente β_1 en el modelo?

En el modelo de regresión binomial con liga log-log los coeficientes $\beta_i \forall i \in [1, 5]$ implican la determinación de la probabilidad de éxito en la distribución binomial de la siguiente manera:

$$\eta_i = \beta_1 + \beta_2 x_{50i} + \beta_3 x_{100i} + \beta_4 x_{200i} + \beta_5 x_{500i}$$

Lo que varía, respecto a la liga logit, es la definición de p_i con base en el predictor lineal se tiene para el modelo log-log:

$$p_i = \frac{\exp(\exp(\eta_i)) - 1}{\exp(\exp(\eta_i))}$$

El modelo para la Y_i sigue siendo binomial por lo que:

$$\Rightarrow Y_i \sim \text{Bin}(p_i, C_i)$$

En este sentido nota que si $x_{50_i} = x_{100_i} = x_{200_i} = x_{500_i} = 0$ entonces se tiene $n_i = \beta_1$, por lo que:

$$\hat{Y}_i = E(Y_i|p_i, c_i, X_i) = E(Y_i|p_i, c_i, x_{20_i}) = \beta_{1,i}C_i \text{ donde } i \in [1, 12]$$

Y, \hat{Y}_i denota la cantidad esperada (promedio) de billetes de 20 pesos falsos como resultado de una cantidad total de billetes de 20 pesos emitidos. En realidad el estimador como esta planteado sólo tiene sentido para el rango de valores donde $i \in [1, 12]$ pues para los restante $i - \text{valores}$ la variable C_i esta representando la cantidad de billetes emitidos de otras denominaciones(en ese caso: $\hat{Y}_i = (\beta_1 + \beta_j)C_i$).

En el caso de la denominación de 20 pesos, se tiene (JAGS):

$$\beta_1 = -6.2$$

$$\Rightarrow \eta_i = \beta_1 = -6.2, \forall i \in [1, 12] \equiv \eta_{20}$$

$$\Rightarrow p_{20} = \frac{\exp(\exp(-6.2)) - 1}{\exp(\exp(-6.2))}$$

```
beta1<-out.sum.jags[grep("beta",rownames(out.sum.jags)),c(1)][1]
p20<-c((exp(exp(beta1))-1)/exp(exp(beta1)))%>%as.data.frame()
colnames(p20)[1]<-c("Prob de Billetes Falsos")
```

```
knitr::kable(p20, caption = "Estimación de la probabilidad de billetes de 20 mxn falsos liga: clog-log"
```

Table 13: Estimación de la probabilidad de billetes de 20 mxn falsos
liga: clog-log

Prob de Billetes Falsos
0.0019812

Con el objetivo de tener una mejor ejemplificación del siguiente inciso a continuación calculo p_{50} :

Se tiene (JAGS):

$$\beta_2 = 2.9$$

$$\Rightarrow \eta_i = \beta_1 + \beta_2 = -6.2 + 2.9 \approx -3.3, \forall i \in [13, 24] \equiv \eta_{50}$$

$$\Rightarrow p_{50} = \frac{\exp(\exp(-3.3)) - 1}{\exp(\exp(-3.3))}$$

```
beta1<-out.sum.jags[grep("beta",rownames(out.sum.jags)),c(1)][1]
beta2<-out.sum.jags[grep("beta",rownames(out.sum.jags)),c(1)][2]
p50<-c((exp(exp(beta1+beta2))-1)/exp(exp(beta1+beta2)))%>%as.data.frame()
```

```
colnames(p50)[1]<-c("Prob de Billetes Falsos")
```

```
knitr::kable(p50, caption = "Estimación de la probabilidad de billetes de 50 mxn falsos liga: clog-log")
```

Table 14: Estimación de la probabilidad de billetes de 50 mxn falsos
liga: clog-log

Prob de Billetes Falsos
0.0375158

f.2

En el modelo de regresión binomial con liga logística, ¿cómo interpretas la suma $\beta_i + \beta_j$ para $j = 2, \dots, 5$?

La respuesta es que corresponde al **cambio del logaritmo asociado a la probabilidad de NO tener billetes falsos, al incluir la denominación j**. Se demuestra de la siguiente manera:

Sea

$$\log(-\log(1 - p_i)) = \begin{cases} \beta_1 & \text{si } i \in [1, 12] \\ \beta_1 + \beta_2 & \text{si } i \in [13, 24] \\ \beta_1 + \beta_3 & \text{si } i \in [25, 36] \\ \beta_1 + \beta_4 & \text{si } i \in [37, 48] \\ \beta_1 + \beta_5 & \text{si } i \in [49, 60] \end{cases}$$

Definamos:

$$i = \{20\} \text{ y } j \in \{\{50\}, \{100\}, \{200\}, \{500\}\}$$

$$\Rightarrow \log(-\log(1 - p_i)) = \beta_i \equiv \beta_{20} \text{ (caso base)}$$

$$\Rightarrow \log(-\log(1 - p_i)) = \beta_1 + \beta_d \text{ donde: } d \in \{\{50\}, \{100\}, \{200\}, \{500\}\}$$

$$\Rightarrow \beta_d = \log(-\log(1 - p_j)) - \log(-\log(1 - p_i))$$

$$\Rightarrow \beta_d = \log \frac{\log(1 - p_j)}{\log(1 - p_i)}$$

$$\Rightarrow e^{\beta_d} = \frac{\log(1 - p_j)}{\log(1 - p_i)} \equiv \frac{\text{logaritmo asociado a la probabilidad de NO tener billetes falsos distintos a 20 mxn}}{\text{logaritmo asociado a la probabilidad de NO tener billetes falsos de 20 mxn}} \equiv \Delta \text{logaritmos c}$$

En particular, recordemos que ya calculamos: $p_{20} \approx 0.001$ y $p_{50} \approx 0.037$ por lo que en el caso de β_{50} , se tiene:

$$\Rightarrow e^{\beta_{50}} = \frac{\log(1 - 0.037)}{\log(1 - 0.001)}$$

```

v1<-p20$`Prob de Billetes Falsos`
v2<-p50$`Prob de Billetes Falsos`
cambio<-c(log(1-v2)/log(1-v1))%>%as.data.frame()
colnames(cambio)[1]<-c("Cambio en Log NO billetes falsos")

knitr::kable(cambio, caption = "Cambio en logaritmos p50: clog-log",row.names = FALSE)

```

Table 15: Cambio en logaritmos p50: clog-log

Cambio en Log NO billetes falsos
19.28103

Inciso (g)

En el modelo de regresión binomial con liga logística define “la tasa de billetes falsos por mil circulando” para cada denominación, $p_j, j = 1, \dots, 5$ y estimalas al 95%

NOTA que estas nuevas tasas no se definen igual que en el caso de la liga logística. Encuentra tú la definición correcta para la liga que estas usando.

Recordemos que en el caso de la liga clog-log:

$$p_i = \frac{\exp(\exp(\beta_1 + \beta_j) - 1)}{\exp(\beta_1 + \beta_j)}$$

Por lo que, en este caso, “la tas de billetes falsos por mil circulando” para cada denominación p_j , toma la forma :

$$p_j = 1000 \frac{\exp(\exp(\beta_1 + \beta_j) - 1)}{\exp(\beta_1 + \beta_j)}$$

```

EstimacionP1000_cloglog<-out.sum.jags[grep("p",rownames(out.sum.jags)),c(1,3,7)]%>%as.data.frame()
EstimacionP1000_cloglog$mean<-EstimacionP1000_cloglog$mean*1000
EstimacionP1000_cloglog$`2.5%`<-EstimacionP1000_cloglog$`2.5%`*1000
EstimacionP1000_cloglog$`97.5%`<-EstimacionP1000_cloglog$`97.5%`*1000
knitr::kable(EstimacionP1000_cloglog, caption = "Estimación de la tasa de billetes falsos por mil circulan",row.names = FALSE)

```

Table 16: Estimación de la tasa de billetes falsos por mil circulando
liga: clog-log

	mean	2.5%	97.5%
p[1]	1.996158	1.502042	2.471086
p[2]	1.996158	1.502042	2.471086
p[3]	1.996158	1.502042	2.471086
p[4]	1.996158	1.502042	2.471086
p[5]	1.996158	1.502042	2.471086
p[6]	1.996158	1.502042	2.471086
p[7]	1.996158	1.502042	2.471086
p[8]	1.996158	1.502042	2.471086
p[9]	1.996158	1.502042	2.471086
p[10]	1.996158	1.502042	2.471086

	mean	2.5%	97.5%
p[11]	1.996158	1.502042	2.471086
p[12]	1.996158	1.502042	2.471086
p[13]	37.526325	35.770909	39.309834
p[14]	37.526325	35.770909	39.309834
p[15]	37.526325	35.770909	39.309834
p[16]	37.526325	35.770909	39.309834
p[17]	37.526325	35.770909	39.309834
p[18]	37.526325	35.770909	39.309834
p[19]	37.526325	35.770909	39.309834
p[20]	37.526325	35.770909	39.309834
p[21]	37.526325	35.770909	39.309834
p[22]	37.526325	35.770909	39.309834
p[23]	37.526325	35.770909	39.309834
p[24]	37.526325	35.770909	39.309834
p[25]	23.792765	22.497733	25.094465
p[26]	23.792765	22.497733	25.094465
p[27]	23.792765	22.497733	25.094465
p[28]	23.792765	22.497733	25.094465
p[29]	23.792765	22.497733	25.094465
p[30]	23.792765	22.497733	25.094465
p[31]	23.792765	22.497733	25.094465
p[32]	23.792765	22.497733	25.094465
p[33]	23.792765	22.497733	25.094465
p[34]	23.792765	22.497733	25.094465
p[35]	23.792765	22.497733	25.094465
p[36]	23.792765	22.497733	25.094465
p[37]	16.685587	15.770726	17.612946
p[38]	16.685587	15.770726	17.612946
p[39]	16.685587	15.770726	17.612946
p[40]	16.685587	15.770726	17.612946
p[41]	16.685587	15.770726	17.612946
p[42]	16.685587	15.770726	17.612946
p[43]	16.685587	15.770726	17.612946
p[44]	16.685587	15.770726	17.612946
p[45]	16.685587	15.770726	17.612946
p[46]	16.685587	15.770726	17.612946
p[47]	16.685587	15.770726	17.612946
p[48]	16.685587	15.770726	17.612946
p[49]	14.267451	13.290525	15.319582
p[50]	14.267451	13.290525	15.319582
p[51]	14.267451	13.290525	15.319582
p[52]	14.267451	13.290525	15.319582
p[53]	14.267451	13.290525	15.319582
p[54]	14.267451	13.290525	15.319582
p[55]	14.267451	13.290525	15.319582
p[56]	14.267451	13.290525	15.319582
p[57]	14.267451	13.290525	15.319582
p[58]	14.267451	13.290525	15.319582
p[59]	14.267451	13.290525	15.319582
p[60]	14.267451	13.290525	15.319582

Inciso (h)

h.1

Compara los modelos de regresión binomial con las dos ligas, logística y complementaria log-log. De acuerdo con sus medidas de ajuste determina cuál de los dos es el mejor.

NOTA.- A partir de este momento, para una exposición más limpia, se presentan únicamente los resultados obtenidos mediante la simulaciones del software JAGS

```
ligalogit<-BondadAjusteLigalogit[2,]%>%mutate(Liga="Logit")
ligacloglog<-BondadAjusteLigacloglog[2,]%>%mutate(Liga="clog-log")
bondadjagsligas<-rbind(ligalogit,ligacloglog)
knitr::kable(bondadjagsligas%>%select(Liga,PseudoR,DIC), caption = "Resumen de Medidas de Bondad de Ajuste")
```

Table 17: Resumen de Medidas de Bondad de Ajuste Liga: Logit vs. clog-log

Liga	PseudoR	DIC
Logit	0.5545932	1296.415
clog-log	0.5547261	1296.415

Podemos ver que ambas ligas dan el mismo DIC. Recordemos, que en la practica la **psuedo R** cuadrada **no sirve para comparar entre modelos**. Por lo que a un mismo nivel de DIC es preferible usar el **modelo que mejor interpretabilidad tenga**. Es por lo anterior, que se elige el modelo con liga **logitya** que cada β_j con $j = 1, \dots, 5$ puede interpretarse como el **cambio en el momio asociado a la probabilidad de billetes falsos** en la denominación j respecto al caso base $\beta_1 \equiv \beta_{20}$.

h.2

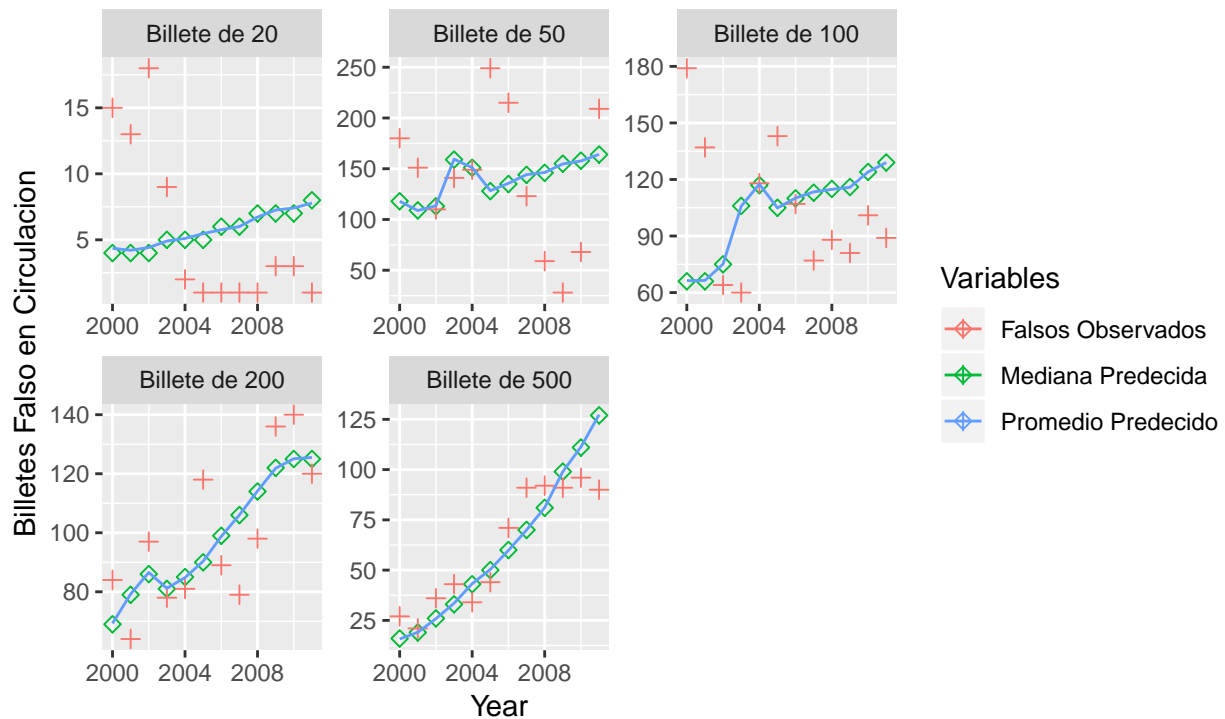
Con el mejor modelo realiza una gráfica de predicción del número de billetes falsos y compáralo con los datos observados. Comenta los puntos importantes de esta gráfica.

```
#Predictions
#cual es el mejor
#CompletoLigaLogit
#CompletoLigacloglog
out.sum.jags<-CompletoLigaLogit
out.yf.best<-out.sum.jags[grepl("yf",rownames(out.sum.jags))],]%>%as.data.frame()
VarResp<-bills_model%>%select(Y)
base<-cbind(out.yf.best,VarResp)%>%cbind(bills_model%>%select(Year))
base$ID<-seq.int(nrow(base))
base$Denominación<-ifelse(base$ID<=12,"Billete de 20",ifelse(base$ID>12 & base$ID<=24,"Billete de 50",i
base$Denominación<-factor(base$Denominación,levels = c("Billete de 20","Billete de 50","Billete de 100"

#ggplot(data=base)+geom_point(aes(x=Year,y=`50%`,colour=Denominación,color="First"),size=2, shape=23)+g
p<-ggplot(data=base)+geom_point(aes(x=Year,y=`50%`,color="Mediana Predecida"),size=2, shape=23)+geom_po
p+scale_fill_brewer(palette = "RdYlBu")+labs(y="Billetes Falso en Circulacion") + ggtitle(expression(at
```

Billetes Falsos en Circulación

Por Denominación



En la grafica anterior observamos que las estimaciones que más se ajustan a los datos observados son para los billetes de 500 mxn. En segundo lugar, de manera visual, la denominación cuya estimación mejor ajusto a los datos observado fue la de los billetes de 200 mxn.

Recordemos que dependiendo de la función de pérdida que utilizemos para consolidar la distribución final en un un estimador puntual se tiene:

- Función de Pérdida Cuadrática: Estimador por la media $\hat{Y} = E(Y)$
- Función de Pérdida absoluta: Estimador por la Mediana $\hat{Y} = Med(Y)$

h.3

En particular comenta sobre los billetes de 20 mxn y de 50 mxn.

```
p<-ggplot(data=base%>%filter(ID<=24))+geom_point(aes(x=Year,y=`50%`,color="Mediana Predecida"),size=2, )
p+scale_fill_brewer(palette = "RdYlBu")+labs(y="Billetes Falso en Circulacion" )+ ggtitle(expression(at
```


Billetes Falsos en Circulación

Baja Denominación



Para los billetes de baja denominación, las estimaciones hechas por el modelo, visualmente, parecen ser las que menos ajustaron los datos observados.

Inciso (i)

Con el mejor modelo, compara las estimaciones de “las tasas de billetes falsos por mil circulando” para las cinco denominaciones. Determina cuales de ellas son estadísticamente diferentes justificando tu respuesta con las estimaciones obtenidas.

```
#EstimacionP1000_cloglog
#EstimacionP1000_logit
EstimacionP1000<-EstimacionP1000_logit
EstimacionP1000$ID<-seq.int(nrow(EstimacionP1000))
EstimacionP1000$Denominación<-ifelse(EstimacionP1000$ID<=12,"Billete de 20",ifelse(EstimacionP1000$ID>12,"Billete de 50",ifelse(EstimacionP1000$ID>24,"Billete de 100",ifelse(EstimacionP1000$ID>36,"Billete de 200","Billete de 500"))))
EstimacionP1000$Denominación<-factor(EstimacionP1000$Denominación,levels = c("Billete de 20","Billete de 50","Billete de 100","Billete de 200","Billete de 500"))
EstimacionP1000<-cbind(EstimacionP1000,bills_model%>%mutate(Promedio=(Y/C)*1000)%>%select(Promedio))

knitr::kable(EstimacionP1000%>%group_by(Denominación)%>%summarise(Promedio=mean(Promedio),Media=mean(Mediana),
2.5=quantile(Promedio,0.025),97.5=quantile(Promedio,0.975)))
```

Table 18: Comparación de tasas de billetes falsos por mil circulando

Denominación	Promedio	Media	2.5%	97.5%
Billete de 20	2.421174	2.011924	1.546752	2.437069
Billete de 50	38.836284	37.512022	35.806281	39.356996
Billete de 100	26.114323	23.783714	22.500079	25.113533
Billete de 200	16.746469	16.676160	15.734583	17.643933

Denominación	Promedio	Media	2.5%	97.5%
Billete de 500	15.792336	14.278264	13.266720	15.326897

Como se puede observar en la tabla 18 ninguno de los intervalos de confianza al 95% de las estimaciones de la tasa de billetes falsos por mil circulando no se intersectan para cualquier par de denominaciones. Por lo que se puede concluir que todas son significativamente distintas entre ellas.

Es decir, nuestro modelo predice un promedio en la tasa de billetes falsos por mil circulando distinto para cada tipo de billete y este efecto es significativamente distinto entre tipo de billete.

En comparación con el **Promedio** observado podemos notar que aunque la **Media** estimada es estadísticamente distinta de 0, para ninguno de los tipos de denominación el intervalo de confianza al 95% capturó el verdadero valor promedio de la tasa de billetes falsos por cada mil observada en los datos.

Como lo observamos al inicio, probablemente esto se deba a la baja **Pseudo R cuadrada** del modelo.

Código BUGS

Modelo Binomial con liga Logit

```
model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbin(p[i],ne[i])
    mu[i]<-ne[i]*p[i]
  }
  #Liga logistica
  # logit(p[i])<-beta[1]+beta[2]*x[i]
  eta[i]<-beta[1]+beta[2]*x1[i]+beta[3]*x2[i]+beta[4]*x3[i]+beta[5]*x4[i]
  p[i]<-exp(eta[i])/(1+exp(eta[i]))
}
#Priors
for (j in 1:5) { beta[j] ~ dnorm(0,0.001) }
#Prediction 1
for (i in 1:n) { yf[i] ~ dbin(p[i],ne[i]) }
}
```

Modelo Binomial con liga clog-log

```
model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbin(p[i],ne[i])
    mu[i]<-ne[i]*p[i]
  }
  #Liga clog-log
  # cloglog(p[i])<-beta[1]+beta[2]*x[i]
  eta[i]<-beta[1]+beta[2]*x1[i]+beta[3]*x2[i]+beta[4]*x3[i]+beta[5]*x4[i]
  p[i]<-(exp(exp(eta[i]))-1)/exp(exp(eta[i]))
}
#Priors
for (j in 1:5) { beta[j] ~ dnorm(0,0.001) }
```

```
#Prediction 1  
for (i in 1:n) { yf[i] ~ dbin(p[i],ne[i]) }  
}
```