

Examen Regresión Avanzada

José Manuel Martínez del Campo González

Marzo 2022

1 Liga Logit

1.1 a) Encuentra explícitamente la LD_{50} función liga elegida

Tenemos que $g(\theta) = \eta$, con g la función liga y $\eta = \alpha + \beta x$ la función de regresión. Para este primer inciso, elegimos la función logit. Entonces, tenemos que:

$$\begin{aligned} g(\theta) &= \text{logit}(\theta) \\ &= \log\left(\frac{\theta}{1-\theta}\right) \\ &= \eta \\ &= \alpha + \beta x \end{aligned} \tag{1}$$

Luego, como para LD_{50} el valor de $\theta = 0.5$, entonces al sustituir en la función nos queda que:

$$\begin{aligned} g(0.5) &= \text{logit}(0.5) \\ &= \log\left(\frac{0.5}{1-0.5}\right) \\ &= 0 \end{aligned} \tag{2}$$

Entonces, la ecuación LD_{50} es:

$$\alpha + \beta x = 0 \iff x = -\frac{\alpha}{\beta} \tag{3}$$

1.2 b) Ajuste del modelo de regresión binomial

Se hizo el ajuste de un modelo de regresión binomial, con distribuciones iniciales vagas de los parámetros α y β . Estos fueron los resultados obtenidos.

	LogDose	No.Animales	No.Muertes	Dose
1	-0.86	5	0	0.4231621
2	-0.30	5	1	0.7408182
3	-0.05	5	3	0.9512294
4	0.73	5	5	2.0750806

Figure 1: Estos son los datos muestrales proporcionados. La columna de "Dose" fue calculada para aclarar la dosis que se le dio al animal.

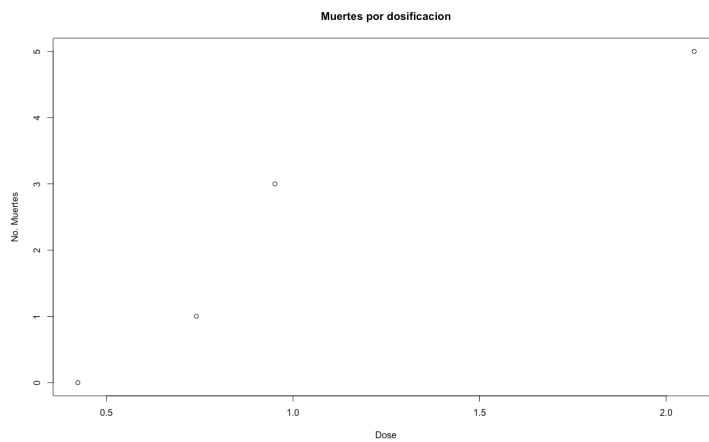


Figure 2: Graficamos la cantidad de dosis que se uso y las muertes.

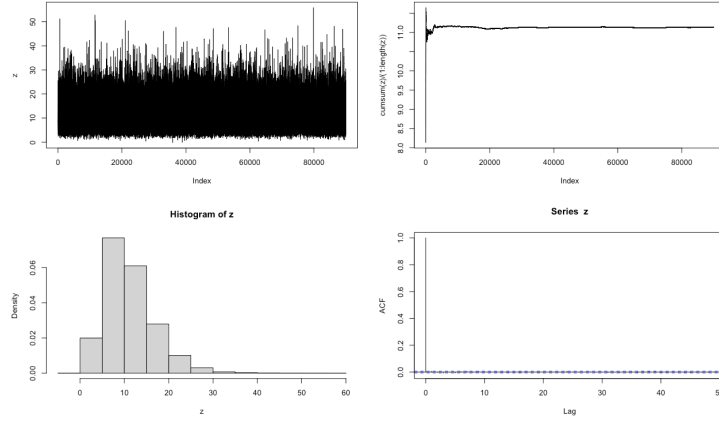


Figure 3: Trace plot, promedios ergónicos, histograma y

	mean	2.5%	97.5%	prob
alpha	1.2612755	-0.6126424	3.6355710	0
beta	11.2251813	3.4119622	24.1220209	0
LD50	-0.1123613	0.1795572	-0.1507159	NaN

Figure 4: Predicciones de los estimadores para el inciso b.

Los datos de la Figura 1 los pasamos al modelo en jags a través de data. Para los inits, vamos a darle los valores de α , β y xf , que es lo que queremos estimar.

En las gráficas de la Figura 3 encontramos los promedios ergónicos del trace plot. Estos nos indican que β va a tender aproximadamente a 11.5. Esto vemos que coincide con la media que calculamos con el modelo para β , lo que se ve bien. El histograma nos muestra la densidad del estimador. Para las series de tiempo, vemos que ...

De la siguiente figura podemos observar que los parametros α y β tienen una correlacion positiva.

El $DIC = 8.303$. Por si solo, este valor no nos dice nada. Veremos otros modelos para comprarlo, y entre menor sea con respecto a estos, mejor.

Por ultimo, para la interpretacion del valor LD_{50} , vemos que su media es $\overline{LD_{50}} = -0.112$. Esto significa que esta cantidad de dosis administrada en escala log nos dará la mediana de los valores, segun la muestra observada.

	mean	2.5%	97.5%	prob
alpha	1.2845118	-0.5860957	3.6589230	0
beta	11.3827836	3.3755703	24.5065536	0
LD50	-0.1128469	0.1736286	-0.1493039	NaN

Figure 5: Predicciones de los estimadores para el inciso c.

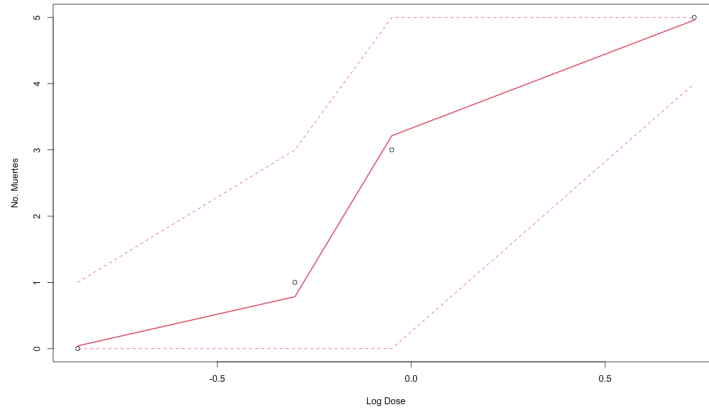


Figure 6: Grafica de las predicciones para el modelo del inciso c).

1.3 c) Ajuste del modelo de regresión binomial

Para este apartado, utilizamos los valores iniciales para los parámetros α , β dados por $E(\alpha) = -17.31$, $V(\alpha) = 1053.72$ y $E(\beta) = 2.57$, $V(\beta) = 23.24$. Con este conocimiento inicial, haremos las predicciones con usando el modelo anterior.

Por otra parte, este modelo obtuvo un $DIC = 8.228$.

1.4 d) Comparacion de los dos modelos y las predicciones

Podemos ver que el mejor modelo es el segundo. Tiene un DIC menor, aun que la diferencia es pequeña. Esto también nos dice que, para estos datos, no es tan relevante tener informacion previa de α y β , ya que la predicción con las iniciales vagas nos dio estimadores muy parecidos.

	mean	2.5%	97.5%	prob
alpha	0.32139427	-1.0118677	1.58370730	0
beta	8.82099043	2.4396376	20.25687566	0
LD50	-0.03643517	0.4147615	-0.07818122	NaN

Figure 7: Predicciones de los estimadores para el inciso f. Inicial vaga.

	mean	2.5%	97.5%	prob
alpha	0.07526696	-1.1413657	1.1910177	0
beta	6.04058116	2.0882236	11.7734125	0
LD50	-0.01246022	0.5465726	-0.1011616	NaN

Figure 8: Predicciones de los estimadores para el inciso f. Inicial informativa.

2 Liga cLog-log

2.1 e) Encuentra explícitamente la LD_{50} función liga elegida

Tenemos que $g(\theta) = \eta$, con g la función liga y $\eta = \alpha + \beta x$ la función de regresión. Para este inciso, elegimos la función cLog-log, con $\theta = 0.5$. Entonces:

$$\begin{aligned}
g(\theta) &= cLoglog(\theta) \\
&= \log(-\log(1 - \theta)) \\
&= \log(-\log(0.693)) \\
&= \eta \\
&= \alpha + \beta x
\end{aligned} \tag{4}$$

Entonces, la ecuación LD_{50} es:

$$\alpha + \beta x = -0.366 \iff x = -\frac{\alpha + 0.366}{\beta} \tag{5}$$

2.2 f) Analisis con distribuciones vagas y con distribuciones iniciales informativas

El modelo cLoglog, segun los datos estimados, se ajusta mejor a los datos. Tanto los datos obtuvidos en el modelo con informacion incial vaga como los que usamos distribucion inicial informativa mejoraron a lo obtenido en la seccion 1.

Podemos ver que los estimadores cambiaron mucho esta vez. Tambien, los DIC que se calcularon con estos dos modelos mejoraron. Para el primero, con

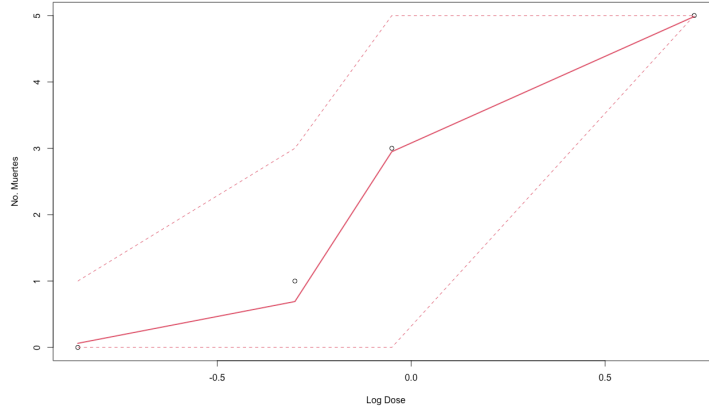


Figure 9: Grafica con las predicciones para el inciso f. Inicial vaga.

informacion vaga, tenemos que $DIC = 8.051$. Para el otro, con la informacion inicial proporcionada, conseguimos un $DIC = 6.770$. Este tuvo una mejora significativa con respecto a los modelos calculados anteriormente.

2.3 h) Resumiendo todos los incisos

Pudimos observar una mejora en el ajuste de los modelos cuando se nos dio informacion inicial. Los primeros tres modelos tuvieron un ajuste parecido. Pudimos ver una clara mejora con el cuarto modelo, el de la liga cLoglog con informativo.

En los modelos se pudo ver que el valor de LD_{50} fue cercano a cero. Esto significa que el valor verdadero de nuestra dosis letal al 50 por ciento sera cercana a 1. En el modelo cLoglog con mejor ajuste, este valor fue de -0.010 . Las graficas de las predicciones no cambiaron mucho de entre las que se hicieron, pero si lo hicieron los estimadores. El ajuste con peor rendimiento fue el logit con inicial no informativa, y el mejor, por lo analizado en los incisos, fue el cLoglog con inicial informativa.

2.4 g) Posterior de LD_{50}

Utilizando el modelo de cLoglog graficamos la posterior de LD_{50} . Podemos ver que, como se habia comentado, el resultado nos dio cercano a 0. Entonces, la dosis que utilizaremos tendra que ser por debajo del 0, en escala logaritmica. En mi opinion, la exposicion a la droga tiene que ser lo suficientemente baja para que no se presenten daños colaterales, intentando maximizar la exposici3n para que los estudios realizados sean efectivos. Es por esto que, un rango por debajo de la dosis letal al 50 por ciento es recomendable, sin embargo que no se

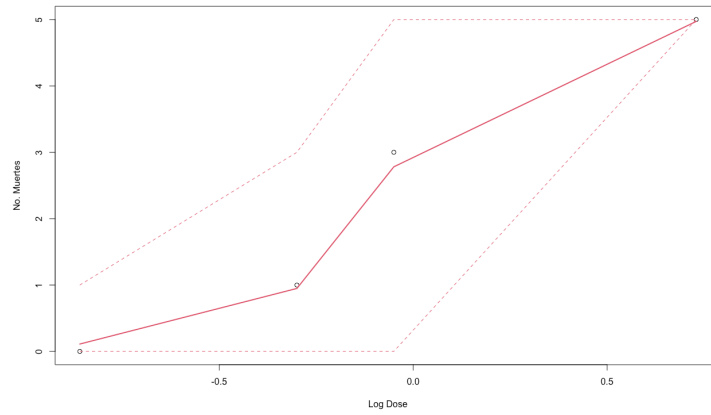


Figure 10: Grafica con las predicciones para el inciso f. Inicial informativa.

baje tanto como para que no sea efectivo el estudio.

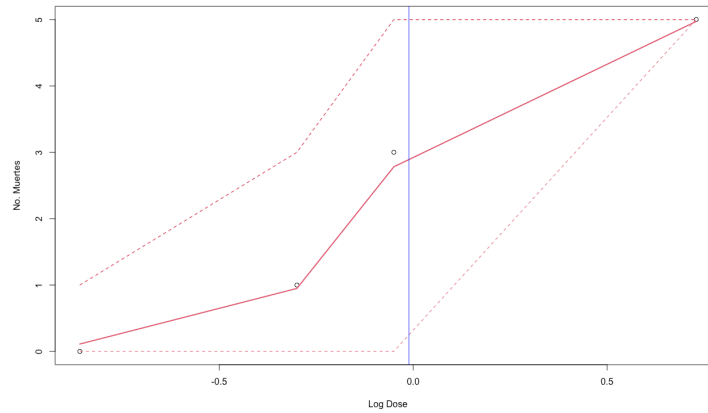


Figure 11: Grafica con las predicciones del modelo cloglog con Inicial informativa. Incluimos la posterior de LD_{50}

3 Apéndice

```
library(R2jags)
library(dplyr)
wdir<-"/Users/joss/Documents/R"
setwd(wdir)

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

mortalidad<-read.csv("http://allman.rhon.itam.mx/~lnieto/index_archivos/Bioassay.csv",header=1)

n <-nrow(mortalidad)
plot(mortalidad)
plot (mortalidad[,1] , mortalidad[,3], xlab = "LogDose", ylab = "No. Muertes", main = "Muertes vs LogDose")
mortalidad <- mutate(mortalidad, Dose = exp(mortalidad$LogDose))
plot (mortalidad[,4] , mortalidad[,3], xlab = "Dose", ylab = "No. Muertes", main = "Muertes vs Dose")

#DATA
nef<-c(4)
xf<-c(4)

data<-list("n"=n,"ne"=mortalidad[,2],"y"=mortalidad[,3],"x"=mortalidad[,1],"nef"=nef,"xf"=xf)
```



```

inits<-function(){list(beta=rep(0,2),yf1=rep(1,n))}

par<-c("beta","p","yf1","yf2")

#Modelo
ej.sim<-jags(data,inits,par,model.file="EjExamen.txt",
             n.iter=50000,n.chains=2,n.burnin=5000,n.thin=1)

#Outs
out<-ej.sim$BUGSoutput$sims.list

z<-out$beta[,2]
par(mfrow=c(2,2))
plot(z,type="l")
plot(cumsum(z)/(1:length(z)),type="l")
hist(z,freq=FALSE)
acf(z)

z<-out$beta
par(mfrow=c(1,1))
plot(z, xlab = "alpha", ylab = "beta")

# Resumen de los estimadores

out.sum<-ej.sim$BUGSoutput$summary

#Tabla resumen
out.sum.t<-out.sum[grepl("beta",rownames(out.sum)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
dimnames(out.sum.t)[[1]][1]<-"alpha"
dimnames(out.sum.t)[[1]][2]<-"beta"
LD50 <- - out.sum.t[1,] / out.sum.t[2,]
out.sum.t <- rbind(out.sum.t, LD50)
dimnames(out.sum.t)[[1]][3]<-"LD50"

print(out.sum.t)

#DIC

out.dic<-ej.sim$BUGSoutput$DIC
print(out.dic)

```

```
### INCISO C ###
```

```
#ReadingDATA
```

```
mortalidad<-read.csv("http://allman.rhon.itam.mx/~lnieto/index_archivos/Bioassay.csv",header
```

```
n <-nrow(mortalidad)
```

```
mortalidad <- mutate(mortalidad, Dose = exp(mortalidad$LogDose))
```

```
#DATA
```

```
xf<-c(4)
```

```
nef<-c(4)
```

```
data<-list("n"=n,"ne"=mortalidad[,2],"y"=mortalidad[,3],"x"=mortalidad[,1],"nef"=nef,"xf"=xf)
```

```
inits<-function(){list(beta=rep(0,2),yf1=rep(1,n))}
```

```
par<-c("beta","p","yf1","yf2")
```

```
#Modelo
```

```
ej.sim<-jags(data,inits,par,model.file="EjExamen.txt",  
             n.iter=50000,n.chains=2,n.burnin=5000,n.thin=1)
```

```
# Resumen de los estimadores
```

```
out.sum<-ej.sim$BUGSoutput$summary
```

```
#Tabla resumen
```

```
out.sum.t<-out.sum[grep("beta",rownames(out.sum)),c(1,3,7)]
```

```
out.sum.t<-cbind(out.sum.t,apply(out$beta,2,prob))
```

```
dimnames(out.sum.t)[[2]][4]<-"prob"
```

```
dimnames(out.sum.t)[[1]][1]<-"alpha"
```

```
dimnames(out.sum.t)[[1]][2]<-"beta"
```

```
LD50 <- - out.sum.t[1,] / out.sum.t[2,]
```

```
out.sum.t <- rbind(out.sum.t, LD50)
```

```
dimnames(out.sum.t)[[1]][3]<-"LD50"
```

```
print(out.sum.t)
```

```
#DIC
```

```
out.dic<-ej.sim$BUGSoutput$DIC
```

```
print(out.dic)
```

```
#Predictions
```

```
out.yf<-out.sum[grep("yf1",rownames(out.sum)),]
```

```
or<-order(mortalidad$LogDose)
```

```

ymin<-min(mortalidad$No.Muertes,out.yf[,c(1,3,7)])
ymax<-max(mortalidad$No.Muertes,out.yf[,c(1,3,7)])

par(mfrow=c(1,1))
plot(mortalidad$LogDose,mortalidad$No.Muertes,ylim=c(ymin,ymax), xlab = "Log Dose", ylab = "No. Muertes", main = "Muertes vs Log Dose")
lines(mortalidad$LogDose[or],out.yf[or,1],lwd=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,3],lty=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,7],lty=2,col=2)

## Inciso D ##

plot (mortalidad[,1] , mortalidad[,3], xlab = "LogDose", ylab = "No. Muertes", main = "Muertes vs Log Dose")
par(new = TRUE)
plot()

## Inciso F ##

#DATA
nef<-c(4)
xf<-c(4)

data<-list("n"=n,"ne"=mortalidad[,2],"y"=mortalidad[,3],"x"=mortalidad[,1],"nef"=nef,"xf"=xf)

inits<-function(){list(beta=rep(0,2),yf1=rep(1,n))}

par<-c("beta","p","yf1","yf2")

#Modelo
ej.sim<-jags(data,inits,par,model.file="EjExamenD1.txt",
             n.iter=50000,n.chains=2,n.burnin=5000,n.thin=1)

# Resumen de los estimadores

out.sum<-ej.sim$BUGSoutput$summary

#Tabla resumen
out.sum.t<-out.sum[grepl("beta",rownames(out.sum)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
dimnames(out.sum.t)[[1]][1]<-"alpha"
dimnames(out.sum.t)[[1]][2]<-"beta"
LD50 <- - out.sum.t[1,] / out.sum.t[2,]
out.sum.t <- rbind(out.sum.t, LD50)
dimnames(out.sum.t)[[1]][3]<-"LD50"

```

```

print(out.sum.t)

#DIC

out.dic<-ej.sim$BUGSoutput$DIC
print(out.dic)

#Predictions
out.yf<-out.sum[grep("yf1",rownames(out.sum)),]
or<-order(mortalidad$LogDose)
ymin<-min(mortalidad$No.Muertes,out.yf[,c(1,3,7)])
ymax<-max(mortalidad$No.Muertes,out.yf[,c(1,3,7)])

par(mfrow=c(1,1))
plot(mortalidad$LogDose,mortalidad$No.Muertes,ylim=c(ymin,ymax), xlab = "Log Dose", ylab = "
lines(mortalidad$LogDose[or],out.yf[or,1],lwd=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,3],lty=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,7],lty=2,col=2)

#Modelo 2
ej.sim<-jags(data,init,par,model.file="EjExamenD2.txt",
             n.iter=50000,n.chains=2,n.burnin=5000,n.thin=1)

# Resumen de los estimadores

out.sum<-ej.sim$BUGSoutput$summary

#Tabla resumen
out.sum.t<-out.sum[grep("beta",rownames(out.sum)),c(1,3,7)]
out.sum.t<-cbind(out.sum.t,apply(out$beta,2,prob))
dimnames(out.sum.t)[[2]][4]<-"prob"
dimnames(out.sum.t)[[1]][1]<-"alpha"
dimnames(out.sum.t)[[1]][2]<-"beta"
LD50 <- - out.sum.t[1,] / out.sum.t[2,]
out.sum.t <- rbind(out.sum.t, LD50)
dimnames(out.sum.t)[[1]][3]<-"LD50"

print(out.sum.t)

#DIC

out.dic<-ej.sim$BUGSoutput$DIC
print(out.dic)
#Predictions
out.yf<-out.sum[grep("yf1",rownames(out.sum)),]

```

```

or<-order(mortalidad$LogDose)
ymin<-min(mortalidad$No.Muertes,out.yf[,c(1,3,7)])
ymax<-max(mortalidad$No.Muertes,out.yf[,c(1,3,7)])

```

```

par(mfrow=c(1,1))
plot(mortalidad$LogDose,mortalidad$No.Muertes,ylim=c(ymin,ymax), xlab = "Log Dose", ylab = "
lines(mortalidad$LogDose[or],out.yf[or,1],lwd=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,3],lty=2,col=2)
lines(mortalidad$LogDose[or],out.yf[or,7],lty=2,col=2)
abline(v = out.sum.t[3,1], col = "blue")

```