

Examen

Olivia Enríquez Rivera

Instrucciones: Realiza un análisis estadístico inferencial en la base de datos Gestacional.csv

Introducción y objetivo

Al ver la descripción de las variables de la base de datos Gestacional.csv, estos datos son interesantes para realizar un estudio de por ejemplo, las causas que determinan el peso de un bebé al nacer o bien, qué factores influyen para la mortalidad del producto al llegar a término el embarazo. Es de esperarse que el peso o la mortalidad de un bebé recién nacido esté altamente relacionado con si lo madre recibió o no monitoreo durante el embarazo, pero una pregunta más interesante que nos podríamos hacer es si influye por ejemplo la escolaridad de la madre, el estado civil, etc. Esto es parte de lo que se desea investigar.

En este trabajo nos centraremos en estudiar el primer caso, es decir, las causas que influyen en el peso del producto. Se pretende detectar qué variables influyen de manera significativa en el peso del bebé. Para ello se utilizará un análisis de regresión bayesiana mediante un modelo lineal generalizado con supuesto de normalidad.

Análisis exploratorio de los datos

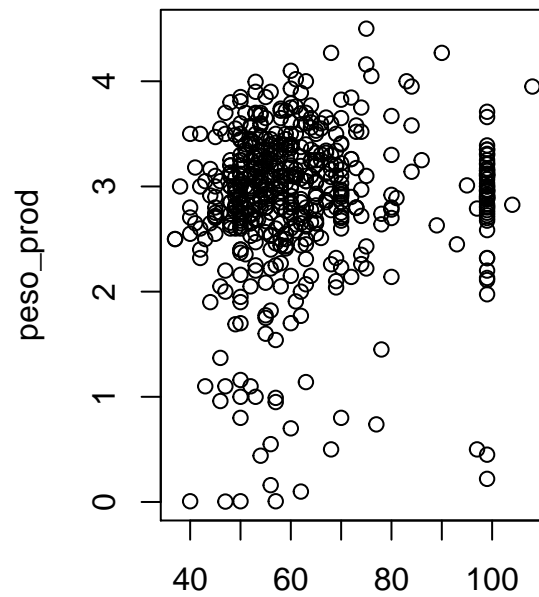
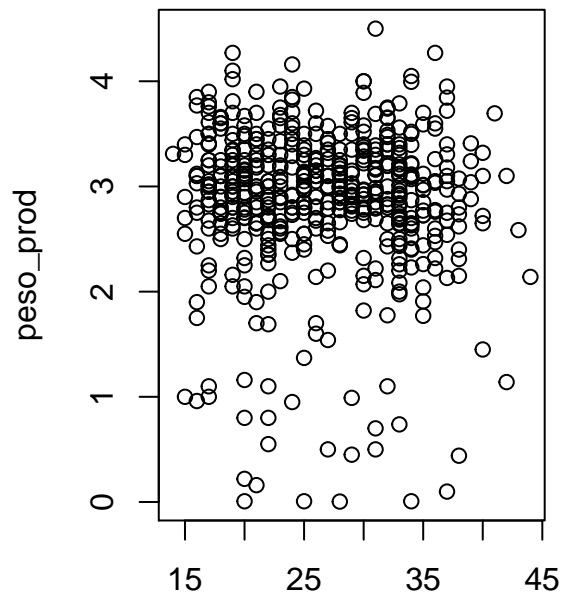
Se trata de 535 casos de embarazo descritos en 12 variables:

1. Anestesia: si se usó durante parto (1-no, 2-sí)
2. Complica: complicación durante el embarazo (1-no, 2-sí)
3. Edad: edad madre en años
4. escolaridad: escolaridad de la madre (1-primaria, 2-secu, 3-prepa, 4-uni)
5. Edo. Civil: edo. civil madre (1-casada, 2-divorciada, 3-soltera, 4-union libre)
6. Forma__nac: tipo parto (1-cesárea, 2-vaginal)
7. Hospital: nombre hospital (1-angeles, 2-español, 3-general, 4-medica sur, 5-oaxaca, 6-pemex sur, 7-tlaxcala)
8. Ind__cesarea: motivo cesarea (1-DCP, 2-electiva, 3-emergencia fetal, 4-emergencia materna, 5-iterativa, 6-no hubo)
9. Monitoreo: si madre recibió monitoreo durante embarzo (1-no, 2-si)
10. Mortalidad: mortalidad del producto (1-no, 2-si)
11. Peso__madre: peso madre inicio embarazo (kgs)
12. Peso__prod: peso al nacer producto (kgs)

Las variables numéricas son: edad, peso__madre, peso__prod

Las variables categóricas son: anestesia, complica, escolaridad, edo. civil, forma__nac, hospital, ind-cesarea, monitoreo, mortalidad.

Observamos en las siguientes gráficas que que al parecer las variables numéricas: el peso de la madre y la edad no están altamente correlacionadas con el peso del producto.



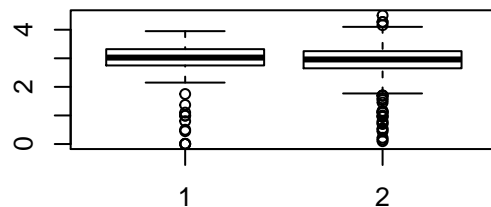
edad

peso_ma

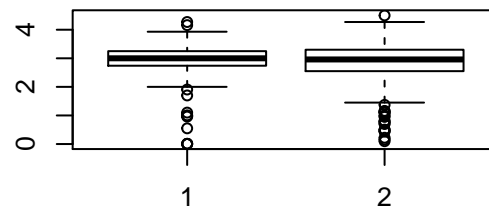
A con-

tinuación se muestra la relación del peso del recién nacido con las variables categóricas de la base de datos. Se observa que el peso del producto tiende a ser mayor cuando: no hay anestesia, no hay complicaciones, escolaridad de la madre es secundaria, parto vaginal, existe monitoreo durante embarazo. Evidentemente, los bebés que llegan a morir tienen muy bajo peso.

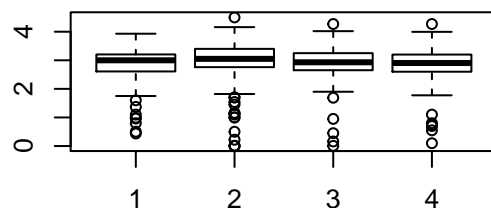
peso prom. por anes



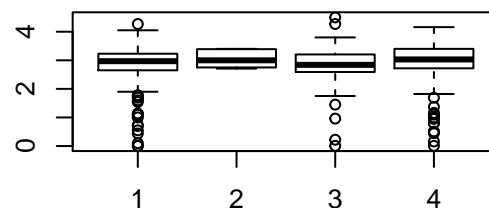
peso prom. por complica



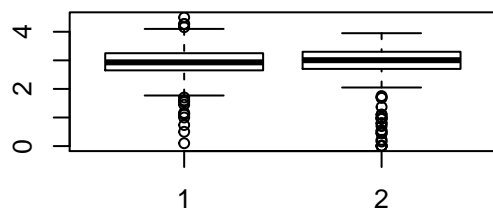
peso prom. por esco



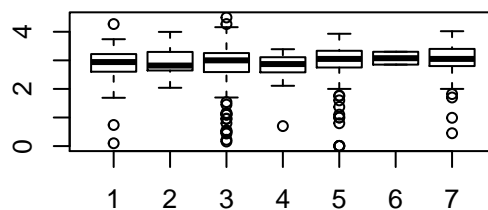
peso prom. por civil



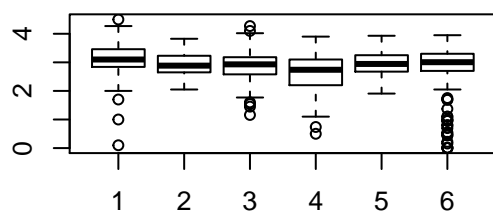
peso prom. por forma



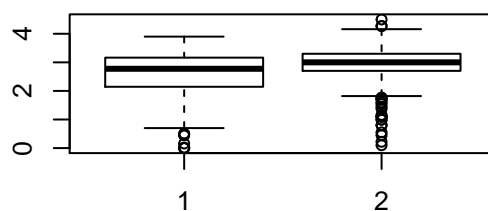
peso prom. por hosp



peso prom. por mot_ces

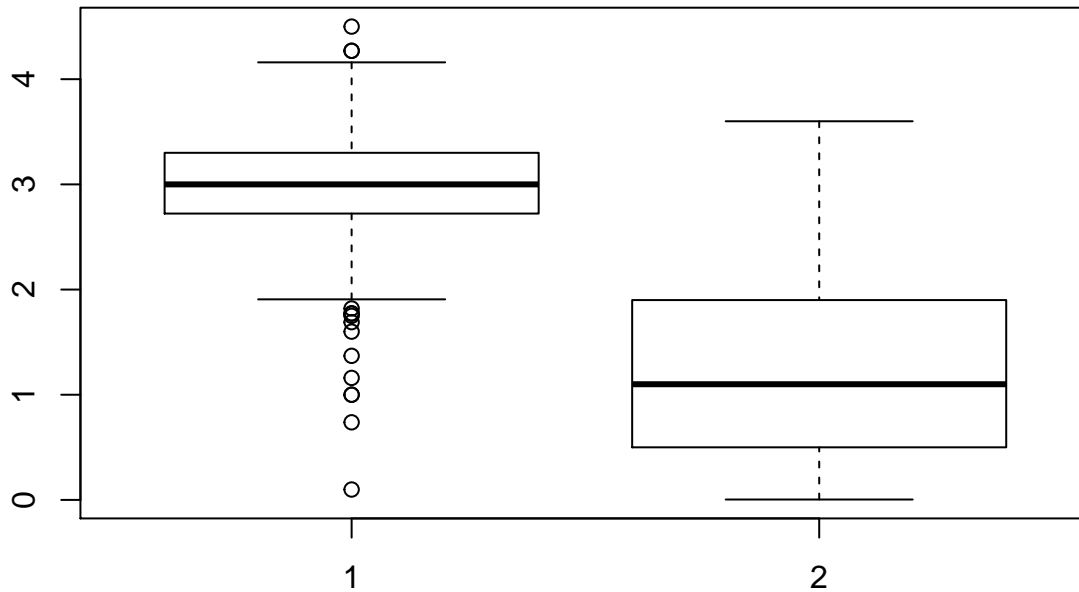


peso prom. por monitor



Otra gráfica boxplot para la variable mortalidad:

peso prom. por mortal



Al parecer, por ejemplo, madres con estado civil soltera tienden a tener hijos con menor peso. Otro dato interesante es que madres del hospital 2: español y 6:pémex sur tienden a tener hijos con buen peso (casi no hay madres de estos hospitales con bajo peso).

Análisis bayesiano

En esta sección se realizará una regresión bayesiana usando un modelo lineal generalizado con supuesto de normalidad. El objetivo es predecir el peso del recién nacido considerando todas las demás variables de la base de datos. El modelo es el siguiente: $Y_{i,j} \sim N(\mu_{i,j}, \tau_i)$

donde μ_{ij} está dado por:

$$\alpha + \beta peso_{ma} + \gamma edad + \delta anes + \epsilon complica + \eta esco + \theta civil + \iota forma + \kappa hosp + \lambda mot_{ces} + \xi monitor + \ni mortal$$

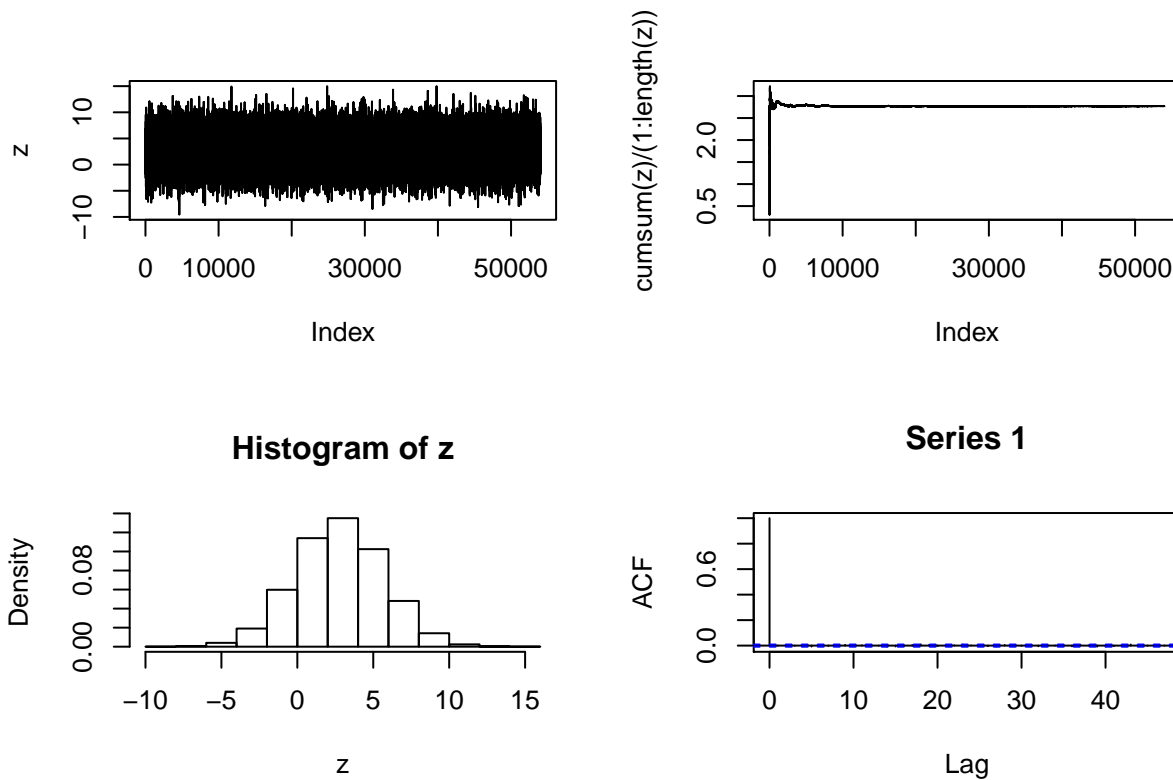
En la primera versión que mandé por correo en la mañana consideré un modelo muy simplificado que contenía únicamente las variables numéricas de la base de datos. En esta segunda versión consideraré a continuación todas las variables de la base de datos, tanto numéricas como categóricas para el modelo.

Como se vio en clase, cuando tratamos variables categóricas debemos imponer restricciones de estimabilidad en el modelo ya que la matriz de diseño se indetermina (existen muchas soluciones para los coeficientes del modelo). El modelo de jags se encuentra en la parte final del reporte. En este modelo las variables aleatorias que terminan con “est” (ver anexo) son las que satisfacen las condiciones de estimabilidad y son las que monitoreamos.

```
## Loading required package: rjags
## Loading required package: coda
## Linked to JAGS 4.3.0
## Loaded modules: basemod,bugs
##
## Attaching package: 'R2jags'
## The following object is masked from 'package:coda':
##
##      traceplot
## module glm loaded
## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 535
##   Unobserved stochastic nodes: 1102
##   Total graph size: 7657
##
## Initializing model
```

En la gráfica que sigue a continuación podemos ver algunos resultados del modelo de jags, en este caso para la variable alfa.est.

En la gráfica se observa a) sup izquierda: la traza de la cadena para el caso de la variable aleatoria alfa, que es el intercepto de nuestro modelo. b)sup. derecha: promedios ergódicos, donde se observa que la cadena alcanzó estabilidad (convergió), c) inferior izquierda: función de densidad para variable aleatoria alfa. d) Gráfica de autocorrelación . si no tiene valores grandes quiere decir que las simulaciones son casi independientes. En este caso, el promedio que se obtuvo para alfa.est es de 2.7763 (kgs), que es el nivel global.



El resumen de los estimadores es el siguiente:

	mean	2.5%	97.5%
## alfa.est	2.771417885	-2.851241e+00	8.378076e+00
## beta	0.002180976	-6.236807e-04	4.961844e-03
## delta.est[1]	-0.050109654	-1.326023e-01	3.678034e-02
## delta.est[2]	0.050109654	-3.678034e-02	1.326023e-01
## deviance	917.512963200	8.341646e+02	1.003984e+03
## epsilon.est[1]	-0.016664023	-6.193689e-02	3.035404e-02
## epsilon.est[2]	0.016664023	-3.035404e-02	6.193689e-02
## etha.est[1]	-0.024005052	-1.250237e-01	7.359932e-02
## etha.est[2]	0.036966477	-4.430778e-02	1.214348e-01
## etha.est[3]	0.036151817	-5.067534e-02	1.200152e-01
## etha.est[4]	-0.049113242	-1.558366e-01	5.519807e-02
## gama	-0.004417552	-1.226203e-02	3.331230e-03
## iota.est[1]	0.033434656	-8.356551e+00	8.434411e+00
## iota.est[2]	-0.033434656	-8.434411e+00	8.356551e+00
## kappa.est[1]	0.003132782	-2.903308e-01	3.075554e-01
## kappa.est[2]	-0.196642046	-4.908494e-01	1.150757e-01
## kappa.est[3]	0.042645724	-2.562646e-01	3.438628e-01
## kappa.est[4]	-0.132669448	-4.311088e-01	1.729256e-01
## kappa.est[5]	0.096925664	-2.093910e-01	4.003782e-01
## kappa.est[6]	0.052479253	-1.684737e+00	1.746267e+00
## kappa.est[7]	0.134128071	-1.722244e-01	4.402721e-01
## lambda.est[1]	0.151640617	-2.645360e+00	2.948412e+00
## lambda.est[2]	0.069071548	-2.731934e+00	2.871071e+00
## lambda.est[3]	-0.124241801	-2.927613e+00	2.673860e+00
## lambda.est[4]	-0.313971155	-3.118971e+00	2.479837e+00
## lambda.est[5]	0.050715202	-2.743191e+00	2.847595e+00
## lambda.est[6]	0.166785589	-1.382226e+01	1.416497e+01

El DIC calculado del modelo 1 es :

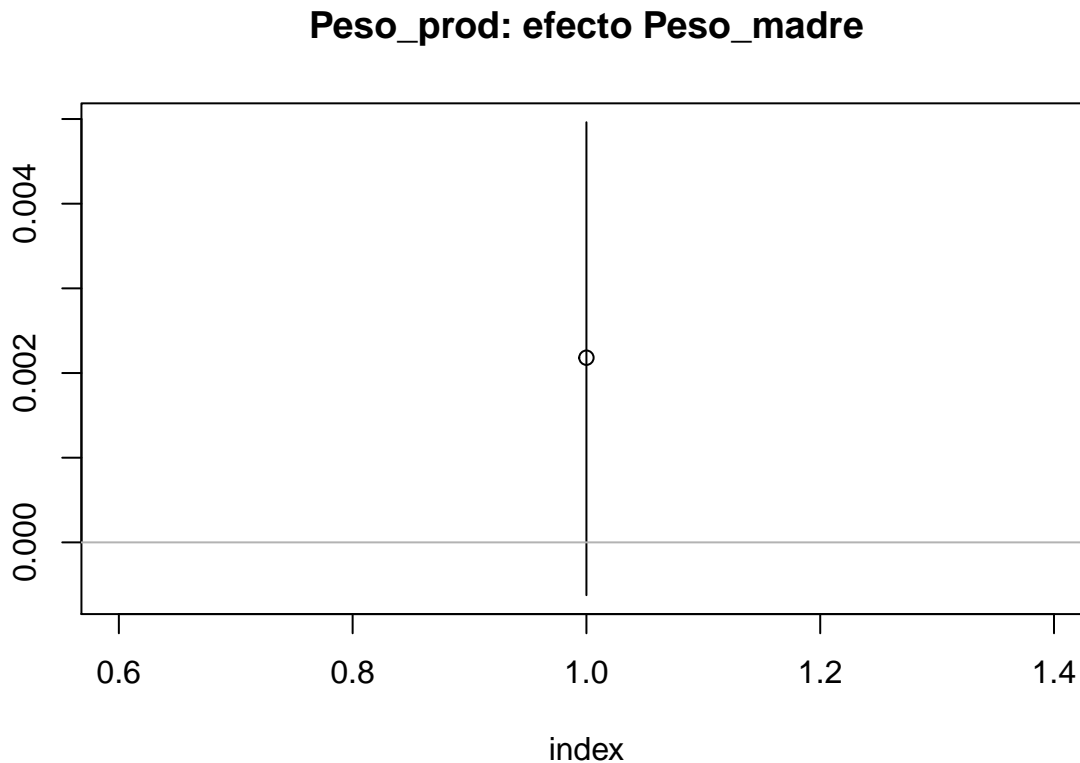
```
## [1] 1957.249
```

Resultados y conclusiones

Coeficiente beta

La media del coeficiente beta es la siguiente para un intervalo de 95% de confianza:

```
##          mean          2.5%          97.5%
## 0.0021809763 -0.0006236807  0.0049618438
```



Vemos en la gráfica anterior que el efecto es positivo, pero como el intervalo contiene al cero, el efecto no es significativo

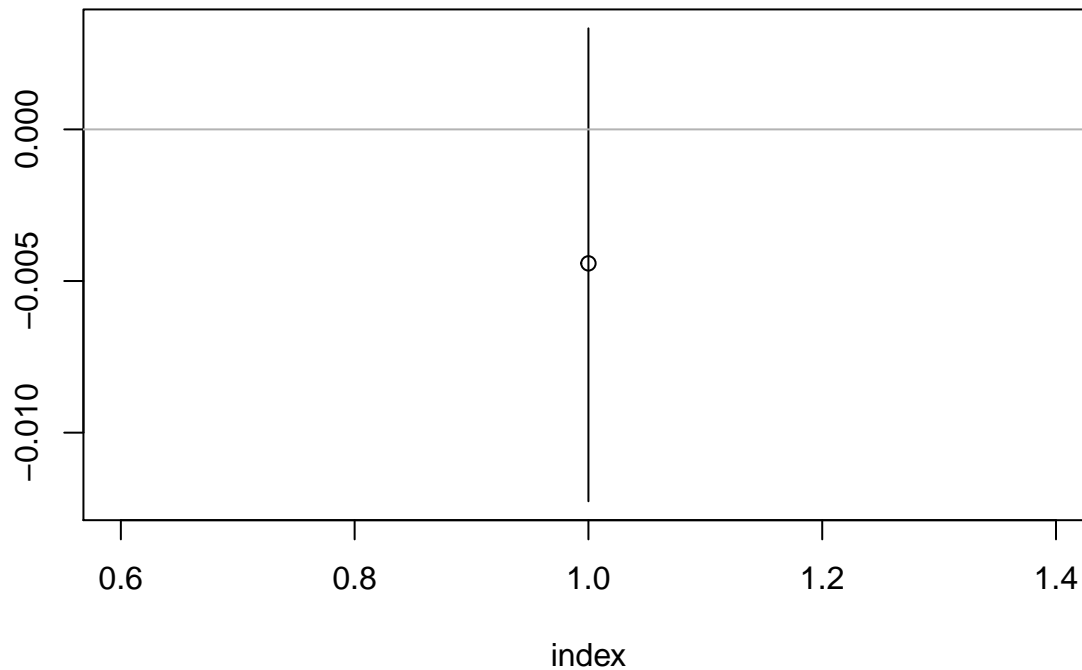
Coeficiente gamma

La media es la siguiente para un intervalo de 95% de confianza:

```
#gama
out.gama<-out.sum[grep("gama",rownames(out.sum)),]
out.est<-out.gama
k<-1
print(out.est[c(1,3,7)])
```

```
##          mean          2.5%          97.5%
## -0.004417552 -0.012262028  0.003331230
```

Peso_prod: efecto edad



Vemos en la gráfica anterior que el efecto es negativo, pero como el intervalo contiene al cero, el efecto no es significativo

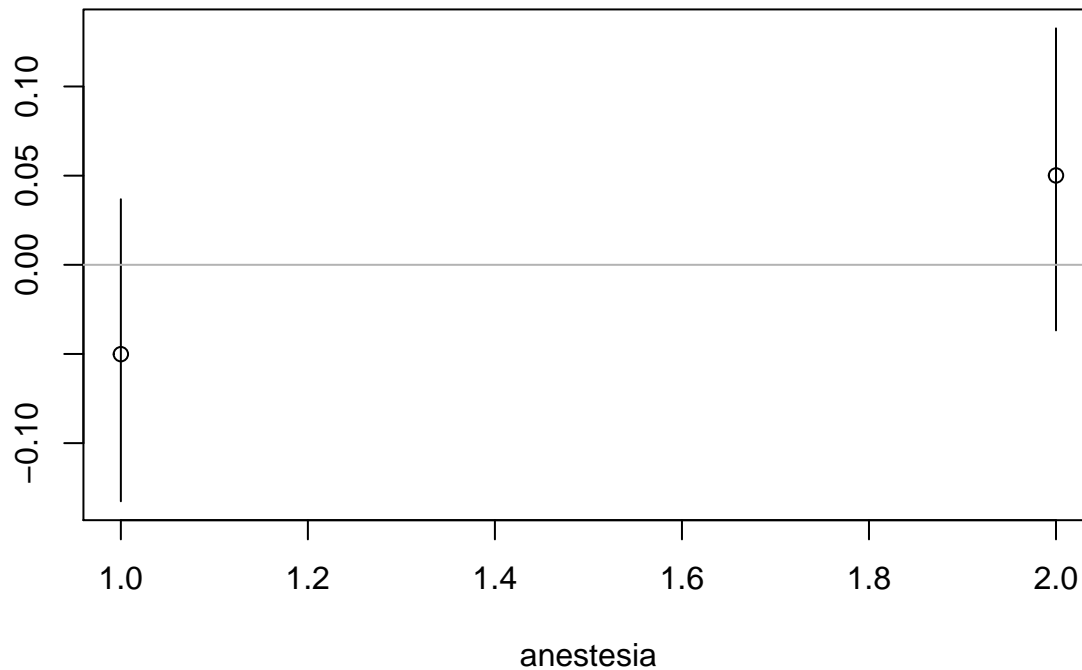
Delta

La media es la siguiente para un intervalo de 95% de confianza:

```
#delta
out.delta<-out.sum[grepl("delta",rownames(out.sum)),]
out.est<-out.delta
k<-2
print(out.est[c(1,3,7)])
```

```
## [1] -0.05010965  0.04332048 -0.07895087
```

Peso_prod: efecto anestesia



Vemos en la gráfica anterior que el efecto es negativo (positivo) para cuando no hay anestesia (si hay anestesia), pero como el intervalo contiene al cero, el efecto no es significativo.

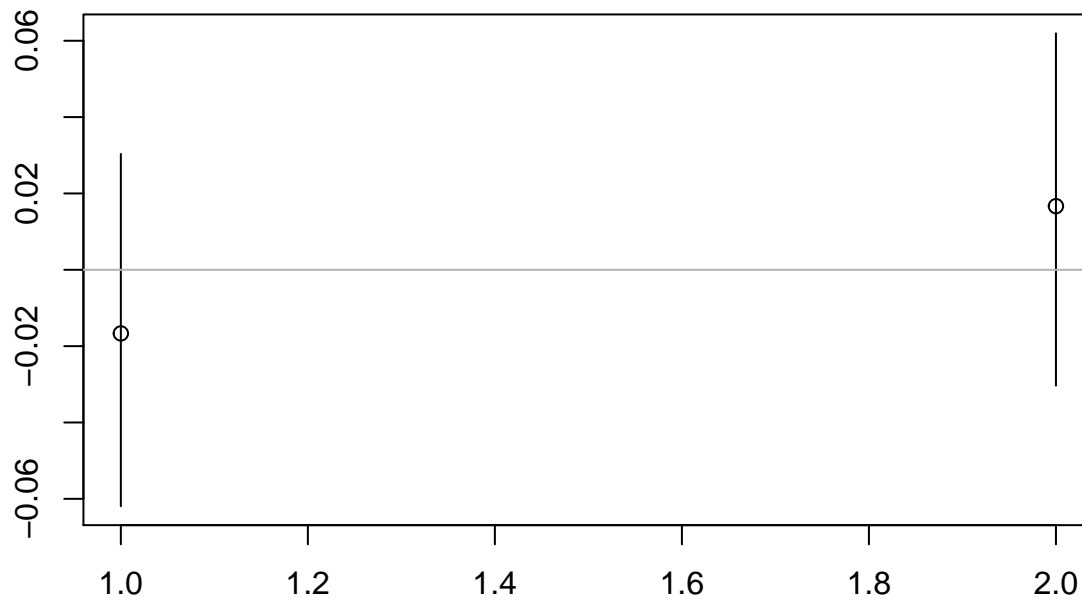
epsilon

La media es la siguiente para un intervalo de 95% de confianza:

```
#epsilon
out.epsilon<-out.sum[grep("epsilon",rownames(out.sum)),]
out.est<-out.epsilon
k<-2
print(out.est[c(1,3,7)])
```

```
## [1] -0.01666402  0.02353086 -0.03257589
```


Peso_prod: efecto complica



complica

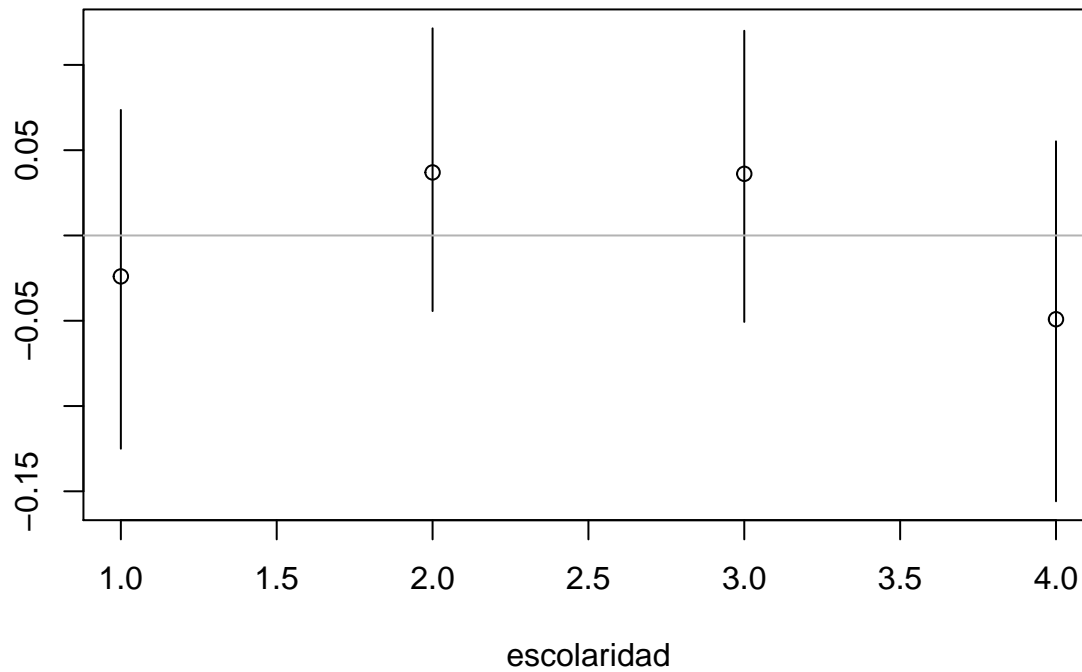
Vemos en la gráfica anterior que el efecto es negativo (positivo) para cuando no hay complicaciones (si hay complicaciones), pero como el intervalo contiene al cero, el efecto no es significativo.

etha

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] -0.02400505  0.03615182  0.04363144
```

Peso_prod: efecto escolaridad



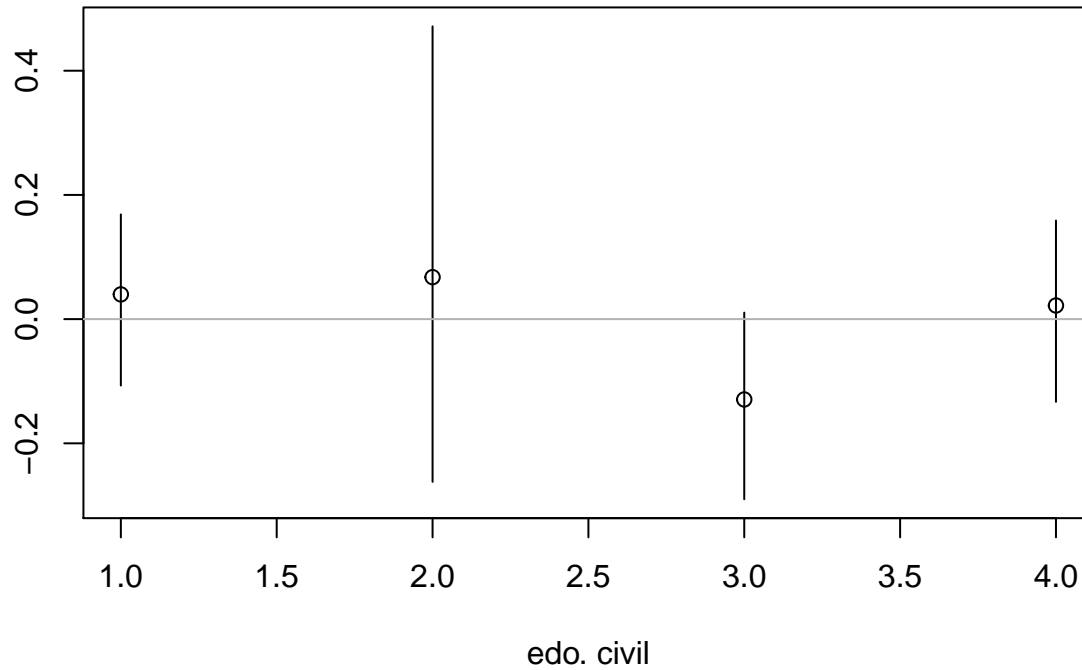
De la gráfica anterior observamos que el efecto de escolaridad 1 (primaria) y 4 (universidad) es negativo en el peso del producto, mientras que para el caso de secundario y preparatoria es positivo. Sin embargo tampoco son significativos.

theta

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] 0.03980816 -0.12935116 0.07905770
```

Peso_prod: efecto edo.civil



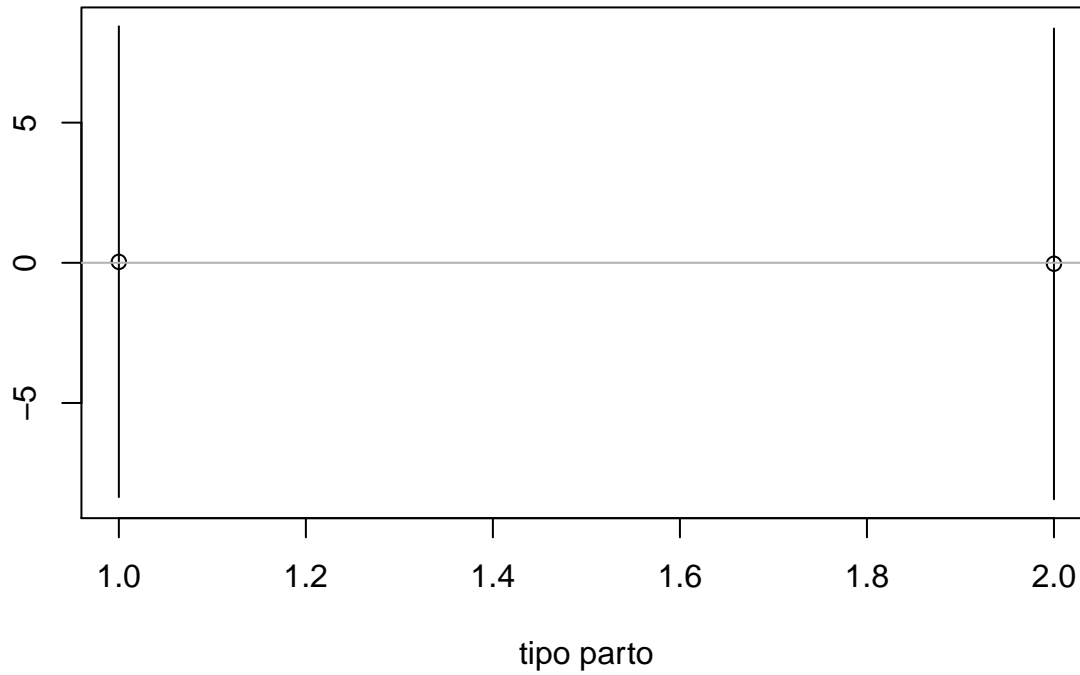
De la gráfica anterior vemos que el efecto de edo. civil 1 (casado) divorciado, union-libre es positivo, mientras que para soltero es negativo, como se había visto en los datos. En todos los casos no es significativo porque tocan la línea del cero.

iota

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] 0.03343466 4.28617877 -2.87204158
```

Peso_prod: efecto tipo parto



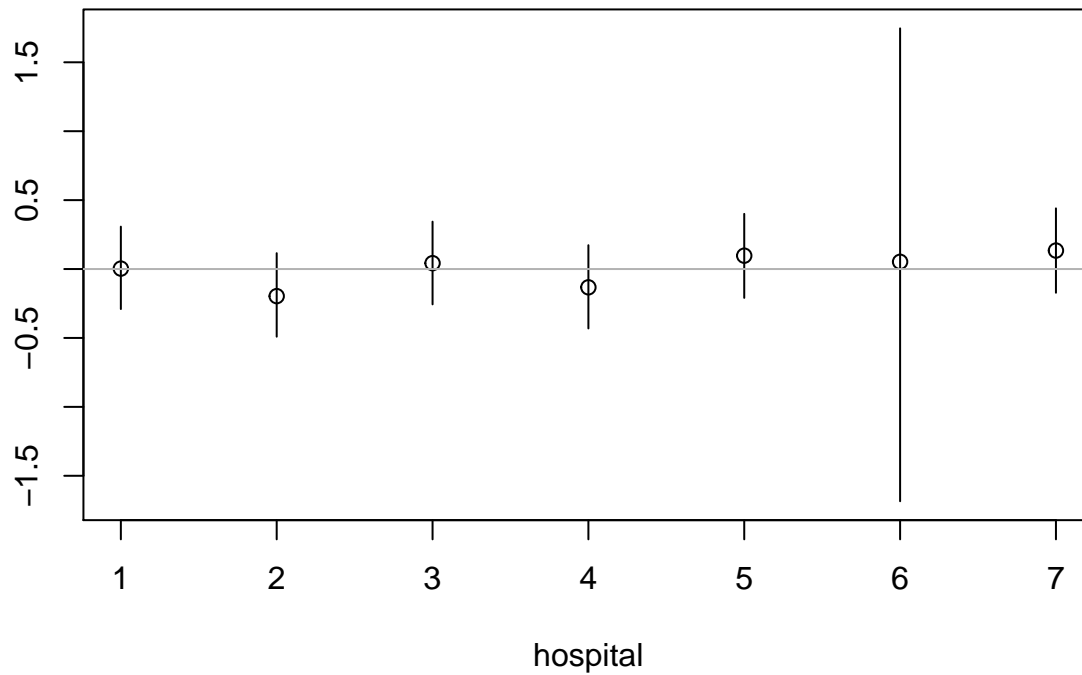
De la gráfica anterior vemos que el efecto tipo parto casi no influye (es muy pequeño el coeficiente) y no es significativo.

kappa

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] 0.003132782 0.042645724 0.134128071
```

Peso_prod: efecto hospital

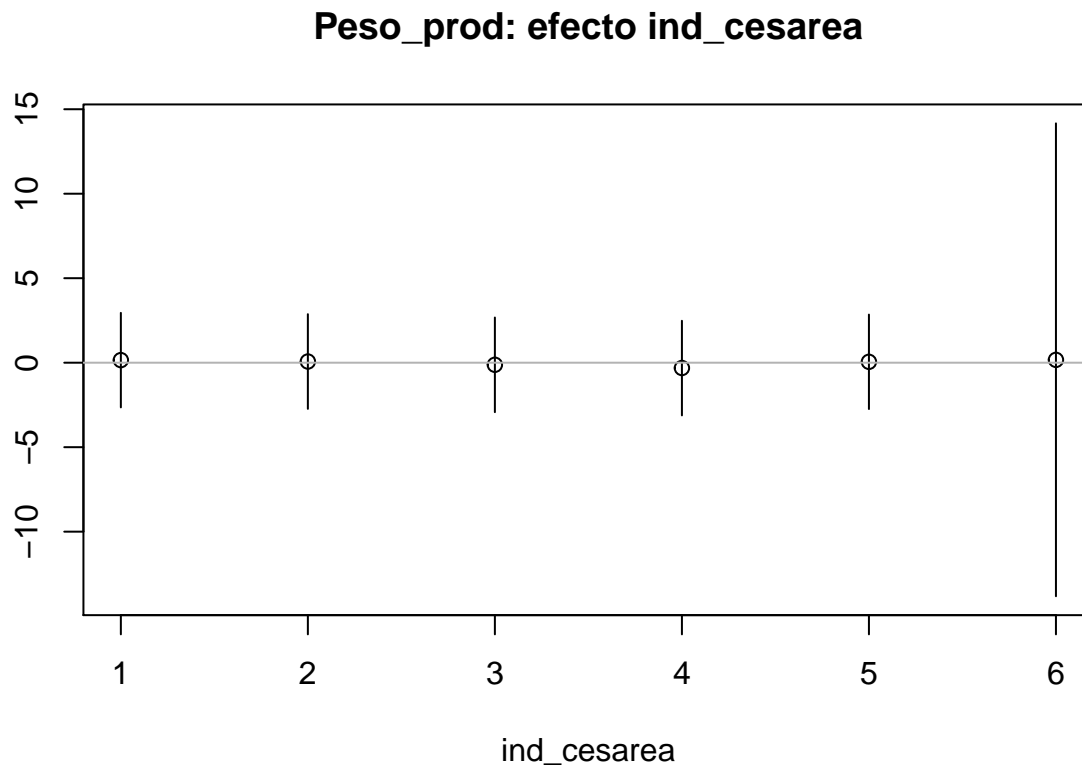


De la gráfica anterior vemos que el tipo de hospital casi no tiene efecto en el peso del recién nacido, ya que es muy pequeño. Influye de manera positiva en los hospitales: 3,5, 6 y 7 que son General, Oaxaca , Pemex sur y Tlaxcala, respectivamente. En los otros influye de manera negativa. En todos los casos no es significativo el efecto.

lambda

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] 0.1516406 -0.1242418 1.4302997
```



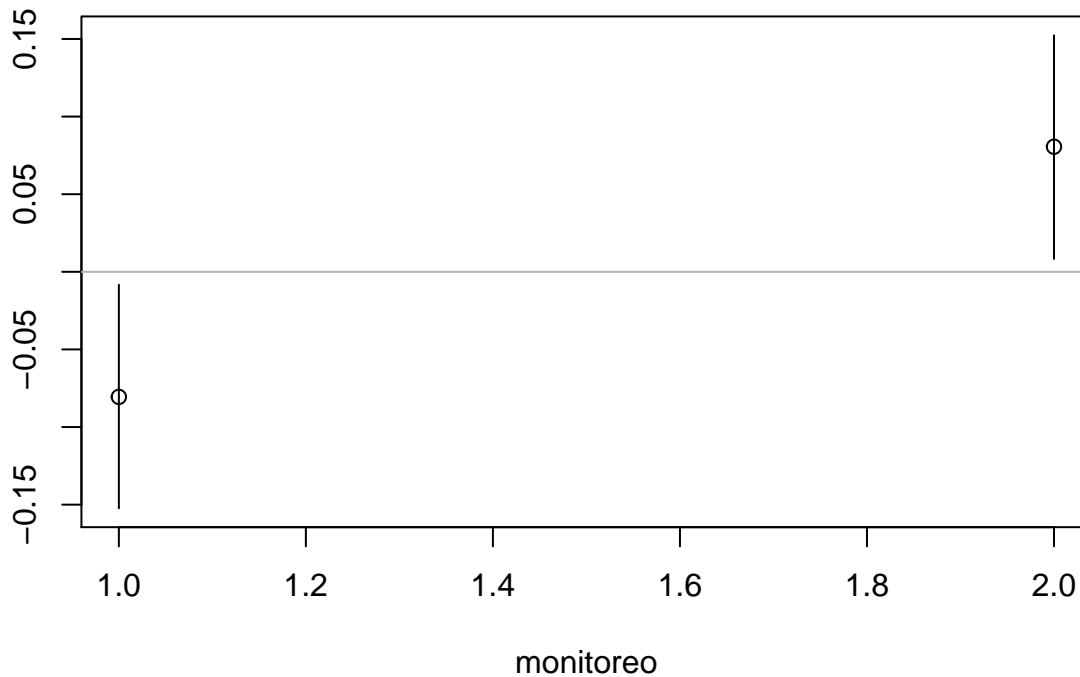
De la gráfica anterior se observa que el efecto ind_cesarea influye casi nada en el peso del recién nacido y no es significativo el efecto.

xi

La media es la siguiente para un intervalo de 95% de confianza:

```
## [1] -0.08057811  0.03832034 -0.10447306
```

Peso_prod: efecto monitoreo



De la gráfica anterior observamos que efecto SI monitoreo influye de manera positiva en el peso del producto. Debido a que el intervalo de confianza no cruza por el cero, concluimos que el efecto es significativo.

Conclusiones

Se investigaron los efectos de todas las variables que se proporcionaron en la base de datos Gestacional.csv, excepto la variable de muerte, en el peso de un bebé recién nacido. El análisis se hizo utilizando regresión bayesiana con un modelo lineal generalizado con supuesto de normalidad. Para hacer el análisis se utilizó el software JAGS (Just Another Gibbs Sampler) y se obtuvieron los coeficientes para las variables involucradas, dando como resultado que la única variable que influye en el peso del recién nacido con un efecto significativo es el monitoreo durante el embarazo. Las demás variables que en el análisis exploratorio parecía tenían algún efecto como: anestesia, complicaciones, escolaridad de la madre, parto vaginal e incluso hospitales y estado civil, no tienen algún efecto significativo en el peso.

ANEXO

Código en R

```
#CODIGO EXAMEN Gestacional.csv
#leemos los datos
datos=read.csv("Gestacional.csv",header=TRUE)
#head(datos)
#numero de datos
n=dim(datos)[1]
#verificamos nulos
f <- function(x){ifelse(x==1,NA,x)}
datos<-lapply(datos, f)
```

```

#cambiamos nombre a las variables
anes<-datos$Anestesia
complica <- datos$Complica
edad <- datos$Edad
esco <- datos$Escolaridad
civil <- datos$Estado_civil
forma <- datos$Forma_nac
hosp <- datos$Hospital
mot_ces <- datos$Ind_cesarea
monitor <- datos$Monitoreo
mortal <- datos$Mortalidad
peso_ma <- datos$Peso_madre
peso_prod <- datos$Peso_prod

#graficamos los datos:
par(mfrow=c(1,2))
#plot(anes,peso_prod)
#plot(complica,peso_prod)
plot(edad,peso_prod)
#plot(esco,peso_prod)
#plot(esco,peso_prod)
#plot(civil,peso_prod)
#plot(forma,peso_prod)
#plot(hosp,peso_prod)
#plot(mot_ces,peso_prod)
#plot(monitor,peso_prod)
#plot(mortal,peso_prod)
plot(peso_ma,peso_prod)

#mas graficas
par(mfrow=c(3,3))
boxplot(peso_prod~anes,main="peso prom. por anes")
boxplot(peso_prod~complica,main="peso prom. por complica")
boxplot(peso_prod~esco,main="peso prom. por esco")
boxplot(peso_prod~civil,main="peso prom. por civil")
boxplot(peso_prod~forma,main="peso prom. por forma")
boxplot(peso_prod~hosp,main="peso prom. por hosp")
boxplot(peso_prod~mot_ces,main="peso prom. por mot_ces")
boxplot(peso_prod~monitor,main="peso prom. por monitor")
boxplot(peso_prod~mortal,main="peso prom. por mortal")

# ANALISIS BAYESIANO: JAGS
library(R2jags)
wdir<-"/Users/olivia/Documents/NIETO/"
setwd(wdir)
#definimos los datos
data<-list("n"=n,"y"=peso_prod, "peso_ma"= peso_ma, "edad"=edad, "anes"=anes, "complica"=complica, "esco"=esco)

#definimos inits
inits<-function(){list(alfa=0, beta=0, gama=0, delta=rep(0,2), epsilon=rep(0,2), etha=rep(0,4), theta=rep(0,4))}

#-Seleccionando parametros a monitorear-
parameters<-c("alfa.est","beta","gama","delta.est","epsilon.est","etha.est","theta.est","iota.est","kappa.est")
#corremos jags

```



```

mod1.sim<-jags(data,inits,parameters,model.file="gestacional_2.txt",n.iter=30000,n.chains=2,n.burnin=3000)
#hacemos trace de las cadenas
#traceplot(mod1.sim,ask=FALSE)

out<-mod1.sim$BUGSoutput$sims.list
z<-out$alfa.est
par(mfrow=c(2,2))
plot(z,type="l")
plot(cumsum(z)/(1:length(z)),type="l")
hist(z,freq=FALSE)
acf(z)
mean(z)

#RESUMEN DE LOS ESTIMADORES

#JAGS
out.sum<-mod1.sim$BUGSoutput$summary
out.sum[1:27,c(1,3,7)]
#write.csv(out.sum[,c(1,3,7)]),file="peso_estim.csv")

#dic
out.dic<-mod1.sim$BUGSoutput$DIC
print(out.dic)

#COEFICIENTES DE LA REGRESION
par(mfrow=c(1,1))

#beta
out.beta<-out.sum[grep("beta",rownames(out.sum)),]
out.est<-out.beta
k<-1
print(out.est[c(1,3,7)])

ymin<-min(out.est[c(1,3,7)])
ymax<-max(out.est[c(1,3,7)])
plot(1:k,out.est[1],xlab="index",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[3],1:k,out.est[7])
abline(h=0,col="grey70")
title("Peso_prod: efecto Peso_madre")

#gama
out.gama<-out.sum[grep("gama",rownames(out.sum)),]
out.est<-out.gama
k<-1
print(out.est[c(1,3,7)])
ymin<-min(out.est[c(1,3,7)])
ymax<-max(out.est[c(1,3,7)])
plot(1:k,out.est[1],xlab="index",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[3],1:k,out.est[7])
abline(h=0,col="grey70")
title("Peso_prod: efecto edad")

#delta
out.delta<-out.sum[grep("delta",rownames(out.sum)),]

```

```

out.est<-out.delta
k<-2
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="anestesia",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto anestesia")

#epsilon
out.epsilon<-out.sum[grepl("epsilon",rownames(out.sum)),]
out.est<-out.epsilon
k<-2
print(out.est[c(1,3,7)])

ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="complica",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto complica")

#eta
out.eta<-out.sum[grepl("eta.est",rownames(out.sum)),]
out.est<-out.eta
k<- 4
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="escolaridad",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto escolaridad")

#theta
out.theta<-out.sum[grepl("theta",rownames(out.sum)),]
out.est<-out.theta
k<-4
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="edo. civil",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto edo.civil")

#iota
out.iota<-out.sum[grepl("iota",rownames(out.sum)),]
out.est<-out.iota
k<-2
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])

```

```

plot(1:k,out.est[,1],xlab="tipo parto",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto tipo parto")

#kappa
out.kappa<-out.sum[grep("kappa",rownames(out.sum)),]
out.est<-out.kappa
k<-7
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="hospital",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto hospital")

#lambda
out.lambda<-out.sum[grep("lambda",rownames(out.sum)),]
out.est<-out.lambda
k<-6
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="ind_cesarea",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto ind_cesarea")

#xi
out.xi<-out.sum[grep("xi",rownames(out.sum)),]
out.est<-out.xi
k<-2
print(out.est[c(1,3,7)])
ymin<-min(out.est[,c(1,3,7)])
ymax<-max(out.est[,c(1,3,7)])
plot(1:k,out.est[,1],xlab="monitoreo",ylab="",ylim=c(ymin,ymax))
segments(1:k,out.est[,3],1:k,out.est[,7])
abline(h=0,col="grey70")
title("Peso_prod: efecto monitoreo")

```

Código en JAGS

```

model
{
  #Likelihood
  for (i in 1:n){
    y[i] ~ dnorm(mu[i],tau[i])
    mu[i] <- alfa + beta*peso_ma[i] + gama*edad[i] + delta[anes[i]] + epsilon[complica[i]] + etha[esco[i]]
  }
  #Priors
  alfa ~ dnorm(0,0.01)
  beta ~ dnorm(0,0.01)

```

```

gama ~ dnorm(0,0.01)
for (i in 1:2){delta[i] ~ dnorm(0,0.01)}
for (i in 1:2){epsilon[i] ~ dnorm(0,0.01)}
for (i in 1:4){etha[i] ~ dnorm(0,0.01)}
for (i in 1:4){theta[i] ~ dnorm(0,0.01)}
for (i in 1:2){iota[i] ~ dnorm(0,0.01)}
for (i in 1:7){kappa[i] ~ dnorm(0,0.01)}
for (i in 1:6){lambda[i] ~ dnorm(0,0.01)}
for (i in 1:2){xi[i] ~ dnorm(0,0.01)}
for (i in 1:n) {tau[i] ~ dgamma(0.01,0.01)}
#Estimability constraints
alfa.est <- alfa + mean(beta[]) + mean(delta[]) + mean(epsilon[]) + mean(etha[]) + mean(theta[]) + mean(iota[]) + mean(kappa[]) + mean(lambda[]) + mean(xi[])
for (i in 1:2){delta.est[i] <- delta[i] - mean(delta[])}
for (i in 1:2){epsilon.est[i] <- epsilon[i] - mean(epsilon[])}
for (i in 1:4){etha.est[i] <- etha[i] - mean(etha[])}
for (i in 1:4){theta.est[i] <- theta[i] - mean(theta[])}
for (i in 1:2){iota.est[i] <- iota[i] - mean(iota[])}
for (i in 1:7){kappa.est[i] <- kappa[i] - mean(kappa[])}
for (i in 1:6){lambda.est[i] <- lambda[i] - mean(lambda[])}
for (i in 1:2){xi.est[i] <- xi[i] - mean(xi[])}
#Prediction
for (i in 1:n) {yf[i] ~ dnorm(mu[i],tau[i])}
}

```