



**Facultad de
Ciencias**
UNAM

ESTADÍSTICA BAYESIANA

Presentación

MODELO JERÁRQUICO NORMAL LINEAL. UN ENFOQUE BAYESIANO
USANDO GIBBS SAMPLER

- Enríquez Hernández Leobardo

12 de noviembre de 2023

Índice

Modelo Jerárquico Normal Lineal. Un enfoque Bayesiano usando Gibbs Sampler aplicado al problema de crecimiento poblacional.	1
Introducción	1
Los datos	1
Metodología	1
Implementación computacional	4
Resultados para el grupo de control.	5
Comparativo con el grupo de tratamiento	14
Conclusiones	16
Referencias	17

Modelo Jerárquico Normal Lineal. Un enfoque Bayesiano usando Gibbs Sampler aplicado al problema de crecimiento poblacional.

Introducción

Esta presentación se basa en (Alan E. Gelfand et al., 1990) en el documento titulado “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling”, en donde los autores discuten el uso del Gibbs Sampler como método para calcular posteriores marginales bayesianas y densidades predictivas, con modelos de datos normales. En todos los casos el enfoque trata de que la especificación distribucional y la implementación computacional sean sencillos. En esta presentación solamente revisaremos la implementación del Gibbs Sampler al modelo jerárquico normal lineal, ya que los autores revisan esto para una variedad de modelos de datos normales, incluidos componentes de varianza, medias ordenadas y desordenadas, y datos faltantes en un ensayo cruzado.

De acuerdo con estos autores la aplicación de los modelos jerárquicos es amplia, sin embargo las metodologías bayesianas para estos modelos son típicamente forzadas a emplear un número de aproximaciones, cuyas consecuencias no son claras bajo las verosimilitudes multiparámetro inducidas por el modelo. En contraste, una implementación completa del enfoque Bayesiano es posible usando Gibbs Sampler, al menos para el modelo normal lineal jerárquico.

Los datos

Para ilustrar el modelo normal lineal jerárquico, los autores se enfocan en un modelo de crecimiento poblacional, basados en un estudio de la empresa CIBA-GEIGY, que midieron el peso de 30 ratas jóvenes en un grupo de control y 30 ratas en un grupo de tratamiento, semanalmente por 5 semanas (x_1, x_2, x_3, x_4 y x_5). El peso de cada una de las ratas fue medido a los 8, 15, 22, 29 y 36 días. Para propósitos de ilustrar la metodología Gibbs Sampler, en una primera parte se enfocan en el grupo de control y no en la comparación con el grupo de tratamiento. Los datos del grupo de control se presentan en el siguiente Cuadro 1.

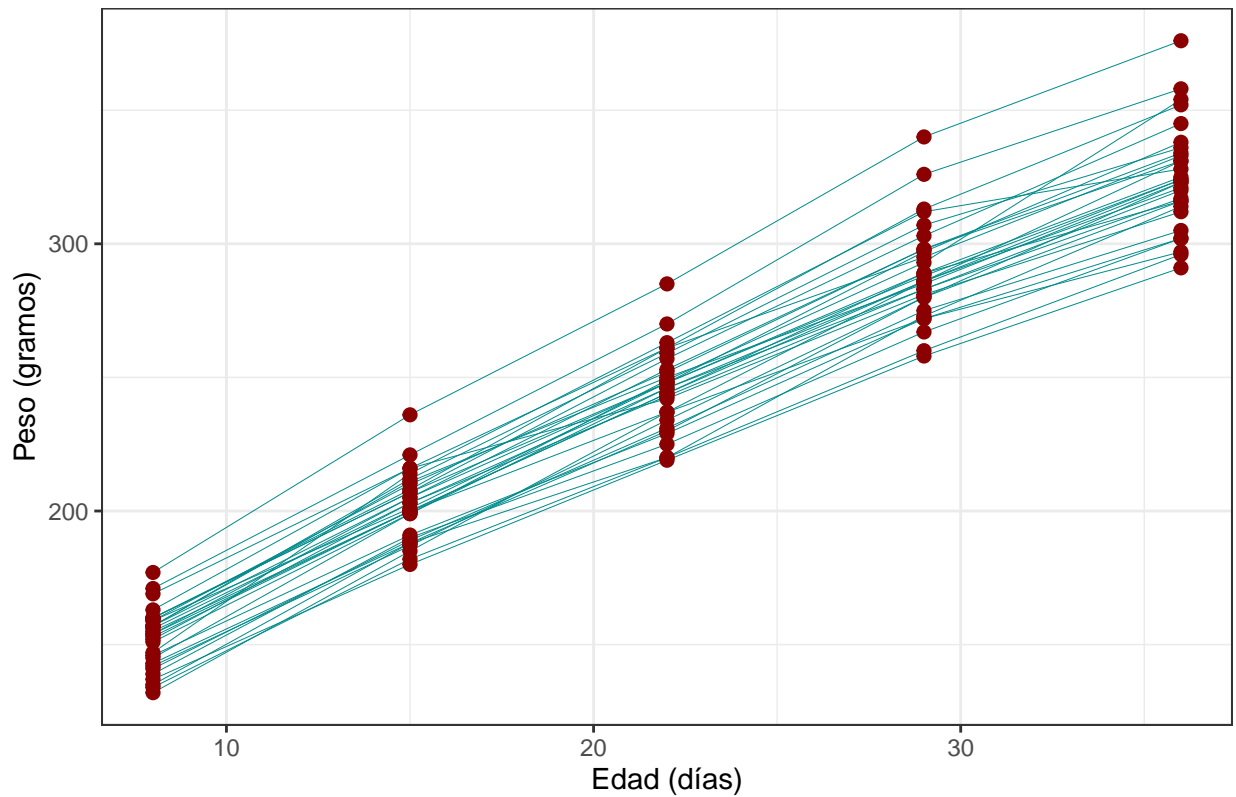
Metodología

Dado el periodo de tiempo considerado argumentan que es razonable asumir curvas de crecimiento individuales en línea recta, como se muestra en la Gráfica 1.

Crecimiento de la población de ratas del grupo de control

Rata	x_1	x_2	x_3	x_4	x_5	Rata	x_1	x_2	x_3	x_4	x_5
1	151	199	246	283	320	16	160	207	248	288	324
2	145	199	249	293	354	17	142	187	234	280	316
3	147	214	263	312	328	18	156	203	243	283	317
4	155	200	237	272	297	19	157	212	259	307	336
5	135	188	230	280	323	20	152	203	246	286	321
6	159	210	252	298	331	21	154	205	253	298	334
7	141	189	231	275	305	22	139	190	225	267	302
8	159	201	248	297	338	23	146	191	229	272	302
9	177	236	285	340	376	24	157	211	250	285	323
10	134	182	220	260	296	25	132	185	237	286	331
11	160	208	261	313	352	26	160	207	257	303	345
12	143	188	220	273	314	27	169	216	261	295	333
13	154	200	244	289	325	28	157	205	248	289	316
14	171	221	270	326	358	29	137	180	219	258	291
15	163	216	242	281	312	30	153	200	244	286	324

Crecimientos individuales



También asumen errores de medición normales homocedásticos (varianzas homogéneas).

Proponen el siguiente modelo de medición jerárquico.

Nivel I. Modelo completo:

$$Y_{ij} \sim N(\beta_{1i} + \beta_{2i}x_{ij}, \sigma^2)$$

donde Y_{ij} y x_{ij} denotan el peso en gramos y la edad en días, respectivamente, de la i -ésima rata al realizarse la j -ésima medición ($i = 1, \dots, 30$ ratas; $j = 1, \dots, 5$ semanas).

Observe que en este caso denotamos la varianza para los datos del grupo de control σ_c^2 (notación de (Alan E. Gelfand et al., 1990)), simplemente como σ^2 . Además, hicimos un cambio de notación para los coeficientes α_i como β_{1i} y β_i como β_{2i} para considerar el vector de estos coeficientes como el “vector de betas”.

Nivel II. La estructura de la población se modela como:

$$\beta_i \sim N_2(\alpha, \Sigma_{\beta*})$$

donde $\beta_i = (\beta_{1i}, \beta_{2i})'$; $i = 1, \dots, 30$.

Observemos que cambiamos la notación original (α_i, β_i) por β_i , y $\mu_c = (\alpha_c, \beta_c)$ de la notación original donde el subíndice c indica que son del grupo de control, por la notación $\alpha = (\beta_1, \beta_2)$.

Por otra parte, la notación original Σ_c se reemplazó por $\Sigma_{\beta*}$.

Nivel III.

$$\alpha \sim N_2(\alpha_0, \Sigma_\alpha),$$

El análisis Bayesiano completo requiere de la especificación a priori para: $\sigma^2 = \sigma_c^2$, $\alpha = \mu_c = (\beta_1, \beta_2)^T$ y $\Sigma_{\beta*} = \Sigma_c$

Se considera el caso en que σ^2 y $\Sigma_{\beta*}$ son desconocidos y se asigna una distribución inicial no informativa a estos parámetros.

El análisis se basará en el método de simulación Gibbs Sampler.

Para la especificación a priori, (Alan E. Gelfand et al., 1990) asumen independencia tomando:

$$[\alpha, \Sigma_{\beta*}^{-1}, \sigma^2] = [\alpha][\Sigma_{\beta*}^{-1}][\sigma^2] \text{ con una forma Normal - Wishart - GamaInversa.}$$

$$\alpha = N(\eta, C), \text{ observe que } \eta = \alpha_0 \text{ y } C = \Sigma_\alpha$$

$$\Sigma_{\beta*}^{-1} = W((\rho R)^{-1}, \rho)$$

$$\sigma^2 = IG(\frac{v_0}{2}, \frac{v_0 \tau_0^2}{2})$$

Reescribiendo el modelo para la i -ésima rata, tenemos:

$$Y_i \sim N(X_i \theta_i, \sigma^2 I_{n_i}) \text{ donde } \theta_i = (\beta_{1i}, \beta_{2i})^T \text{ y } X_i \text{ denota la matriz de diseño apropiada, y definen:}$$

$$Y = (Y_1, \dots, Y_k)^T, \bar{\theta} = k^{-1} \sum_{i=1}^k \theta_i, n = \sum_{i=1}^k n_i, D_i = \sigma^{-2} X_i^T X_i + \Sigma_{\beta*}^{-1}, \text{ y } V = (k \Sigma_{\beta*}^{-1} + C^{-1})^{-1}, \text{ para } i = 1, \dots, k = 30.$$

El Gibbs Sampler para $\theta = (\theta_1, \dots, \theta_k)$, $\Sigma_{\beta*}$ y σ^2 están especificadas para las distribuciones condicionales:

$$[\theta_i | Y, \alpha, \Sigma_{\beta*}^{-1}, \sigma^2] = N\{D_i(\sigma^{-2} X_i^T Y_i + \Sigma_{\beta*}^{-1} \alpha), D_i\}$$

$$[\alpha | Y, \{\theta\}, \Sigma_{\beta*}^{-1}, \sigma^2] = N\{V(k \Sigma_{\beta*}^{-1} \bar{\theta} + C^{-1} \eta, V)\}$$

$$[\Sigma_{\beta*}^{-1} | Y, \{\theta\}, \alpha, \sigma^2] = W\{[\sum_i (\theta_i - \alpha)(\theta_i - \alpha)^T + \rho R]^{-1}, k + \rho\}$$

$$[\sigma^2 | Y, \{\theta\}, \alpha, \Sigma_{\beta*}^{-1}] = IG\{\frac{n+v_0}{2}, \frac{1}{2}[\sum_i (Y_i - X_i \theta_i)^T (Y_i - X_i \theta_i) + v_0 \tau_0^2]\}$$

Para este análisis definen las especificaciones a priori de los hiperparámetros como:

$$C^{-1} = 0, v_0 = 0, \rho = 2, \text{ y } R = \begin{pmatrix} 100 & 0 \\ 0 & 0.1 \end{pmatrix}$$

Lo que refleja una información inicial vaga relativo a la información que proveen los datos.

La simulación de la distribución Wishart para la matriz de 2x2, $\Sigma_{\beta*}^{-1}$, usa el algoritmo de Odell y Feiveson (1996): con $G(.,.)$ distribución Gamma obtenido independientemente de:

$$[U_1] = G(\frac{v}{2}, \frac{1}{2}), [U_2] = G(\frac{v-1}{2}, \frac{1}{2}) \text{ y } [N] = N(0, 1)$$

Estableciendo

$$W = \begin{pmatrix} U_1 & N\sqrt{U_1} \\ N\sqrt{U_1} & U_2 + N^2 \end{pmatrix}$$

Entonces, si $S^{-1} = (H^{1/2})^T (H^{1/2})$

$$\Sigma_{\beta*}^{-1} = (H^{1/2})^T W (H^{1/2}) \sim W(S^{-1}, v)$$

Implementación computacional

Esta implementación se basa principalmente en el modelo presentado por (Alan E. Gelfand et al., 1990), y tomando en cuenta un ejemplo presentado en el libro de (Albert, 2009) para modelos jerárquicos, además de la especificación de la función hiergibbs de la librería de LearnBayes (Albert, May 1, 2019.).

```
set.seed(20241)
hiergibbs_normlin=function(data,m)
{
#####
# Implements Gibbs sampling algorithm for posterior with hierarchical regression prior
#
# INPUT
# data: 30 by 5 matrix where the observed individuals (rats) are
#       in column 1, weights (covariate) are in column 2, in columns 3 and 4
#       we have the week and the day when the sample was taken, in column 5
#       we have the sample size that the row represents.
# m: number of cycles of Gibbs sampling
#
# OUTPUT
# a list with
# -- beta: matrix of simulated values of beta with each row a simulated value
# -- mu: matrix of simulated values of cell means
# -- var: vector of simulated values of second-stage variance sigma^2_pi
#####
y=data[,2]                #vector of Weights (dependent variable)
n=data[,5]                #vector of sample sizes (5weeksx30rats=150=n)
x1=data[,4]               #Age (days), covariates 1 (independent variables)

X=cbind(1+0*x1,x1)        #design matrix
s2=1/5                    #known sample variance sigma^2/n. We assume 1/5
p=2; N=length(y)         #number of parameters beta's

mbeta=array(0,c(m,p))     #sets up arrays to store simulated draws: beta's
mmu=array(0,c(m,length(n))) #sets up arrays to store simulated draws: mu
ms2pi=array(0,c(m,1))     #sets up arrays to store simulated draws: (sigma^2_pi)

##### defines prior parameters
nu=c(107,6) #alpha_0
```

```

C=0
b1=array(nu,c(2,1)) #defines the prior mean for beta's \bar{\beta}

#Wishart distribution for variance and covariance matrix for beta's distrib
#for 2 beta's, m iterations
R=diag(c(100, 0.1), 2, 2)
rho=2
rhoR=rho*R
rhoR_inv=solve(rhoR)
w<-rFractionalWishart(m, rho, rhoR_inv, covariance = FALSE, simplify = "array")
W<-apply(simplify2array(w), 1:2, mean, na.rm = TRUE)#WishartDistrib_simulated
bvar=array(W,c(2,2)) # prior inverse_covariance-variance 2x2 matrix:inverse_SIGMAbeta*
ibvar=solve(bvar) # inverse of inverse_SIGMAbeta*=SIGMAbeta*

s=0; v=0; #hyperparameters of the prior on (sigma^2) with inverse gamma form

#starting values of popu means(mu) and prior variance sigma^2=s2 in Gibbs Samp
mu=y; s2pi=1

for (j in 1:m)
{
## simulates the regression vector beta's,from a multivariate normal distribution,
pvar=solve(ibvar+t(X) %*%X/s2pi) #variance-covariance matrix is stored
pmean=pvar %*%(ibvar %*%b1+t(X) %*%mu/s2pi) #posterior mean (weights) is stored
beta=t(chol(pvar)) %*%array(rnorm(p),c(p,1))+pmean #regression vector beta is stored

#Simulates the prior variance from an inverse gamma distribution
s2pi=(sum((mu-X %*%beta)^2)/2+s/2)/rgamma(1,shape=(N+v)/2)#simulates s2 (sigma^2)

#the components of mu have independent normal distributions: simulates mu
postvar=1/(1/s2+1/s2pi) #values of the posterior variances for the components of mu
postmean=(y/s2+X %*%beta/s2pi)*postvar #posterior means for the components of mu
mu=rnorm(n,postmean,sqrt(postvar)) #simulates values from the n independent norm distr

#stores simulated draws
mbeta[j,]=t(beta) #model beta's
mmu[j,]=t(mu) #post means for weights (dependent variables)
ms2pi[j]=s2pi #model variances sigma^2's
}

return(list(beta=mbeta,mu=mmu,var=ms2pi))
}

m=10000
sol=hiergibbs_normlin(d,m)

```

Resultados para el grupo de control.

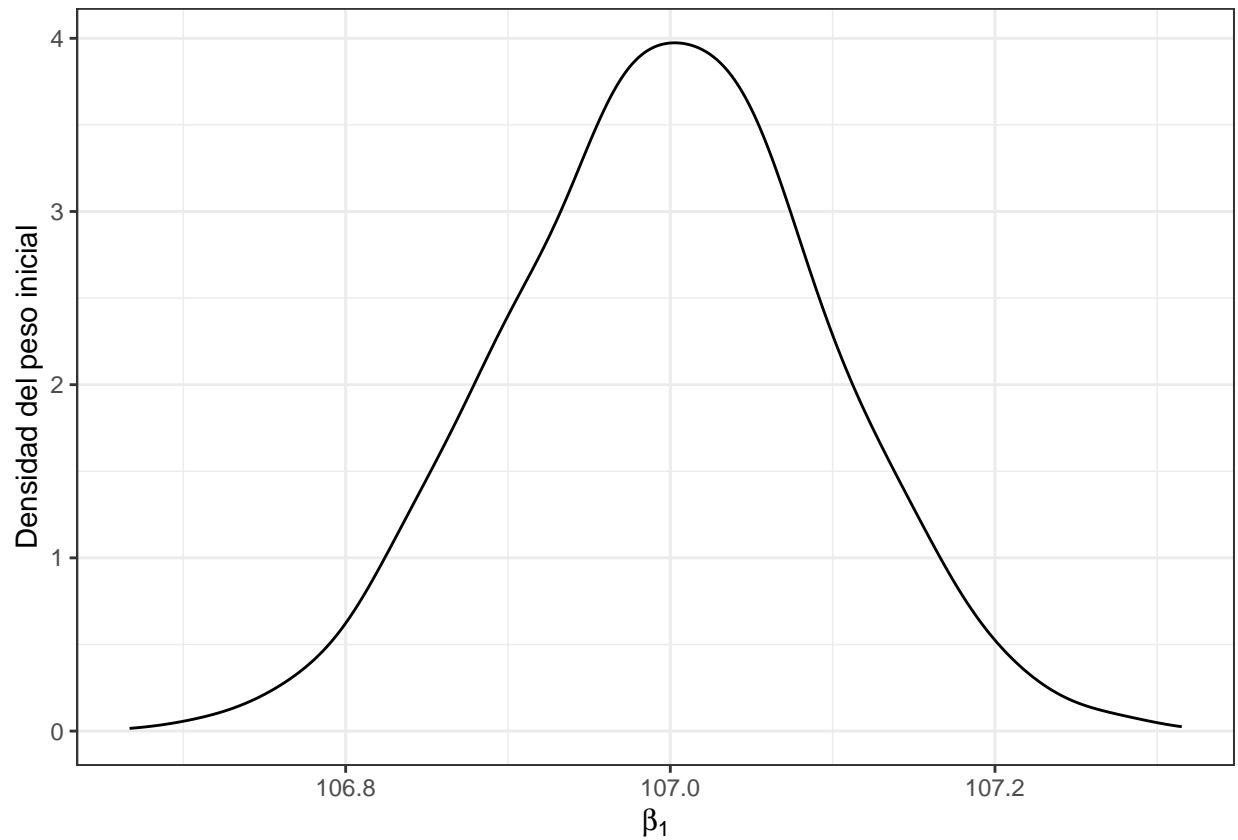
En las siguientes gráficas se presentan las densidades estimadas de las muestras simuladas para los coeficientes de regresión β_1 y β_2 en el modelo de regresión jerárquico.

En la Gráfica 2, tenemos la densidad estimada del peso inicial poblacional.

```

sol_beta1<-sol$beta[-(1:5000),1]
sol_beta1<-as.data.frame(sol_beta1)
p1 <- ggplot(data=sol_beta1, aes(x=sol_beta1)) +
  geom_density(adjust=1.5, alpha=.4)+theme_bw()+
  ylab("Densidad del peso inicial") +
  xlab(expression(beta[1]))
p1

```

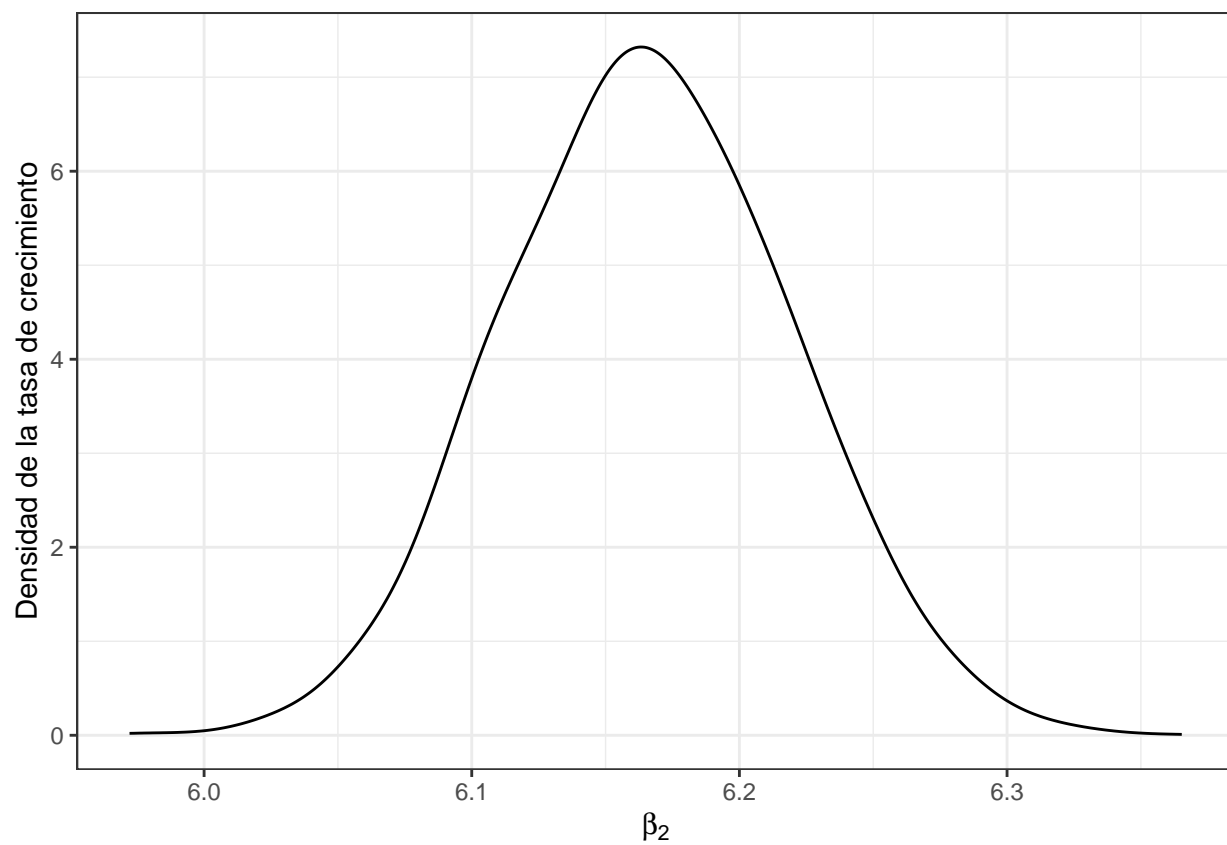


En la Gráfica 3, tenemos la densidad estimada de la tasa de crecimiento.

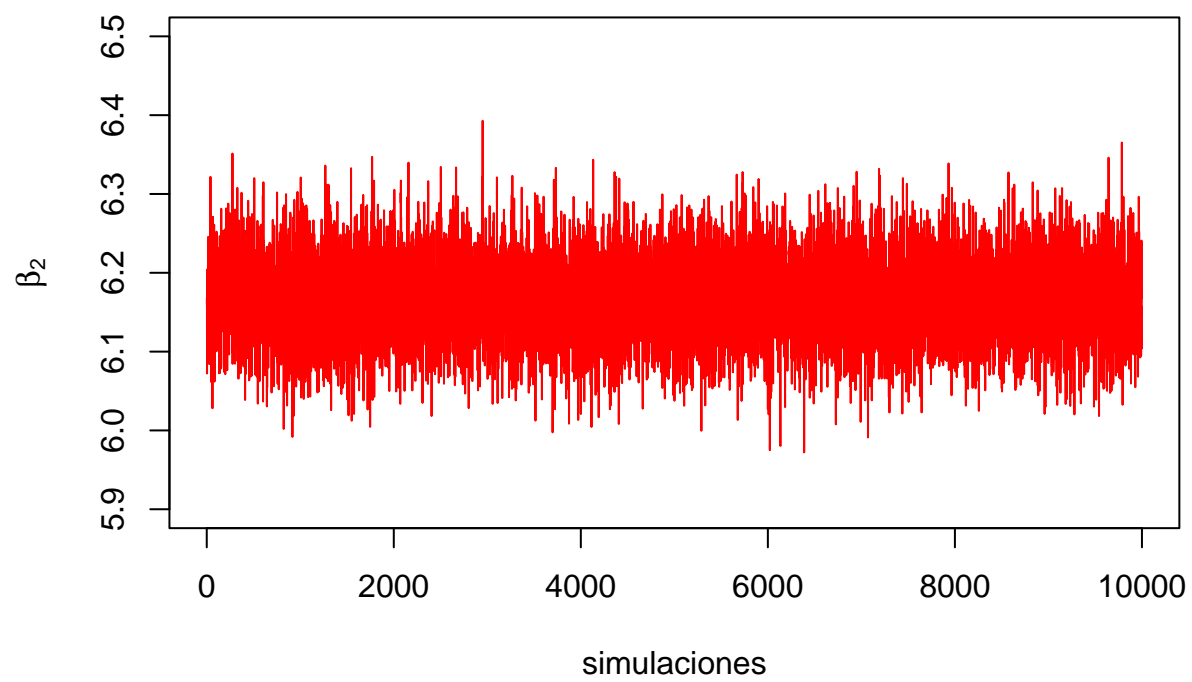
```

sol_beta2<-sol$beta[-(1:5000),2]
sol_beta2<-as.data.frame(sol_beta2)
p2 <- ggplot(data=sol_beta2, aes(x=sol_beta2)) +
  geom_density(adjust=1.5, alpha=.4)+theme_bw() + ylab("Densidad de la tasa de crecimiento") +
  xlab(expression(beta[2]))
p2

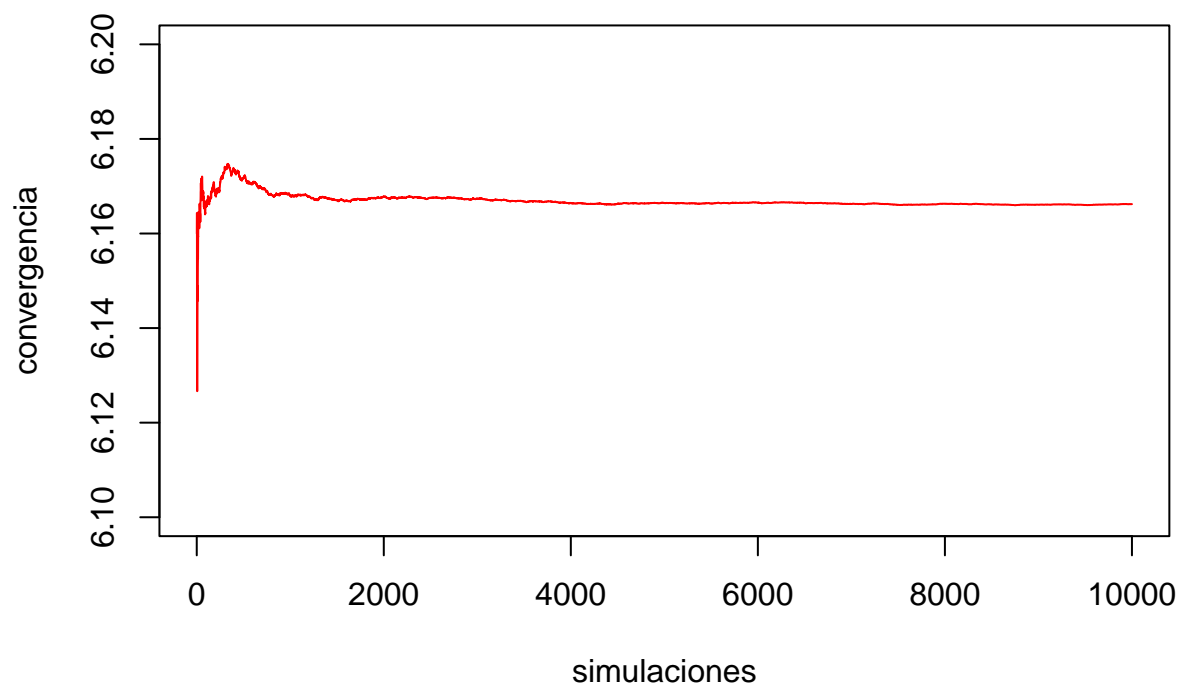
```



En la siguiente Gráfica 4, podemos ver lo siguiente para β_2 :



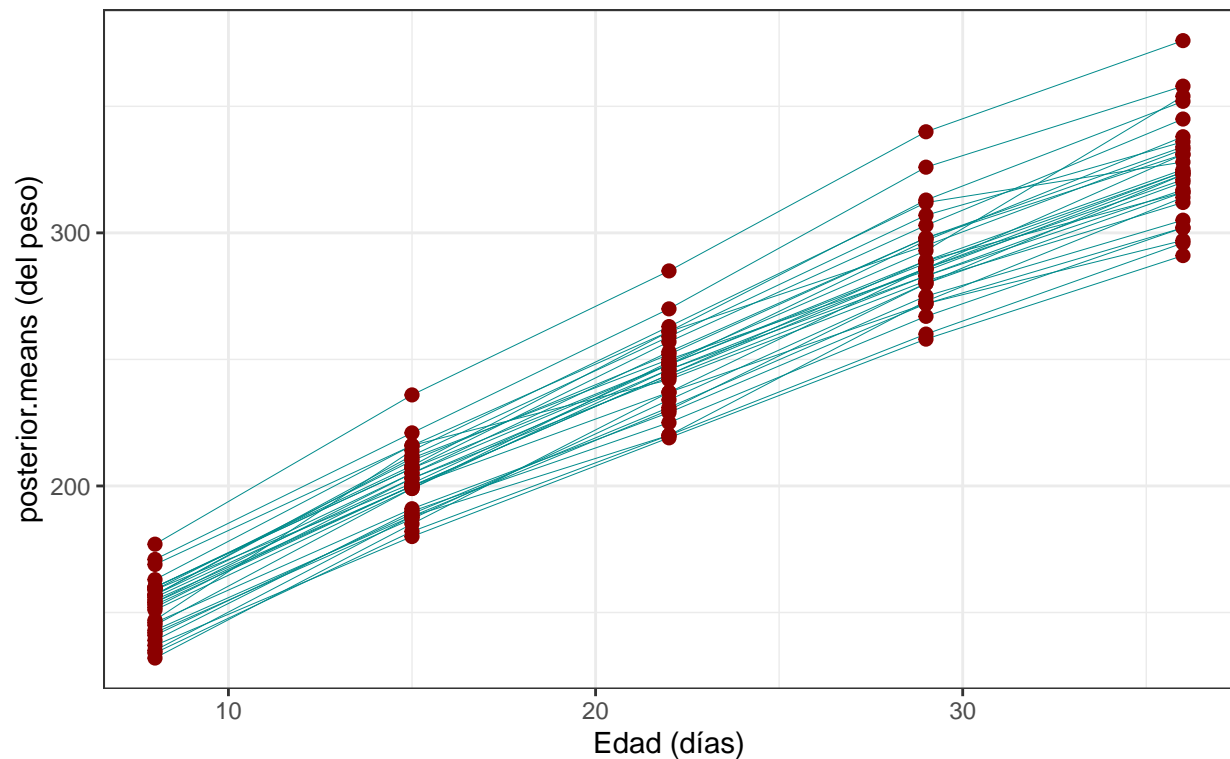
En la Gráfica 5, podemos ver la convergencia para β_2 :



Las medias posteriores para las 30 ratas se presentan en la Gráfica 6.

```
posterior.means = apply(sol$mu, 2, mean)
posterior.means = as.data.frame(posterior.means, nrow = 150, ncol = 1, byrow = T)
d$postmean <- posterior.means$posterior.means
```

Medias posteriores de los pesos individuales



La distribución predictiva final del peso de la Rata 10 a los 45 días de edad se muestra en la siguiente Gráfica 7.

```
d_rat10<-d[46:50,]
m=10000
sol_drat10=hiergibbs_normlin(d_rat10,m)
```

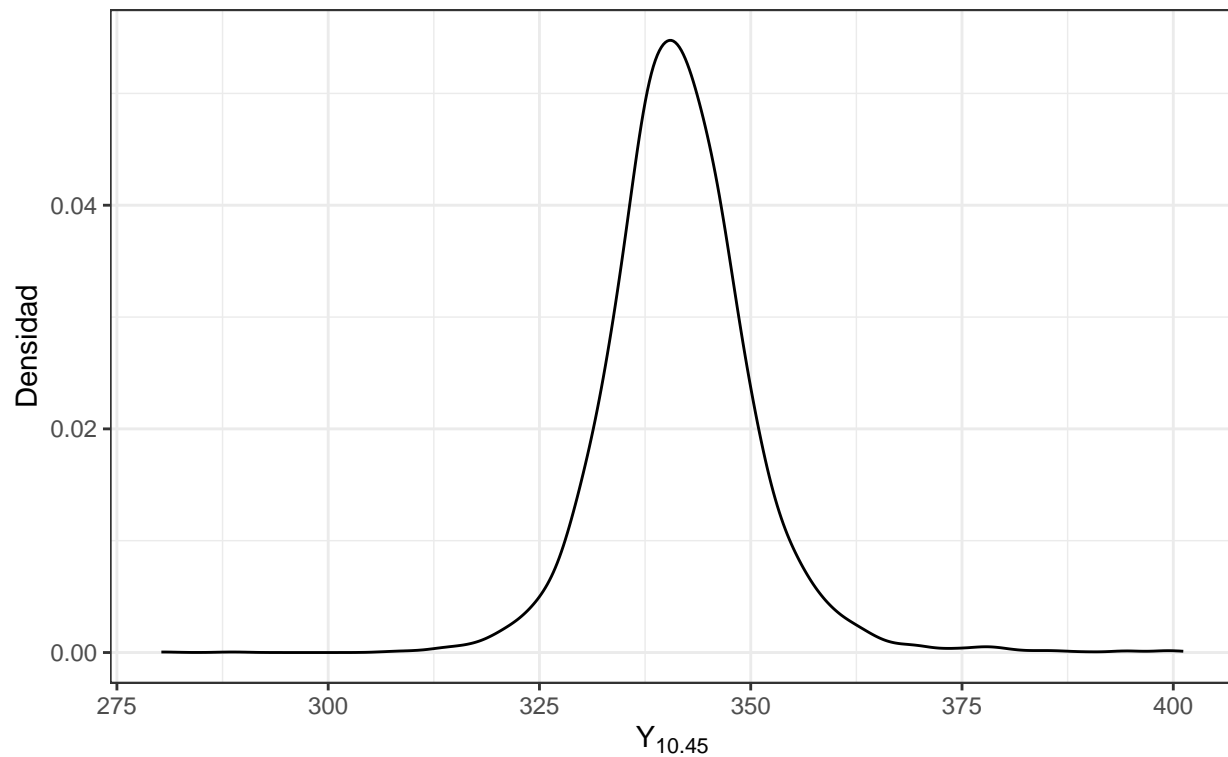
```
rat10<-sol_drat10$mu
rat10<-as.data.frame(rat10)
```

```
prediction<-sol_drat10$beta[,1]+sol_drat10$beta[,2]*45
prediction<-as.data.frame(prediction)
```

```
rat10$d45<-prediction$prediction
```

```
rat10_pred45<-rat10$d45[-(1:5000)]
rat10_pred45<-as.data.frame(rat10_pred45)
p3 <- ggplot(data=rat10_pred45, aes(x=rat10_pred45)) +
  geom_density(adjust=1.5, alpha=.4)+theme_bw() + ylab("Densidad") + ggtitle("Distribución predictiva")
p3
```

Distribución predictiva final del peso de la rata 10 a los 45 días de edad



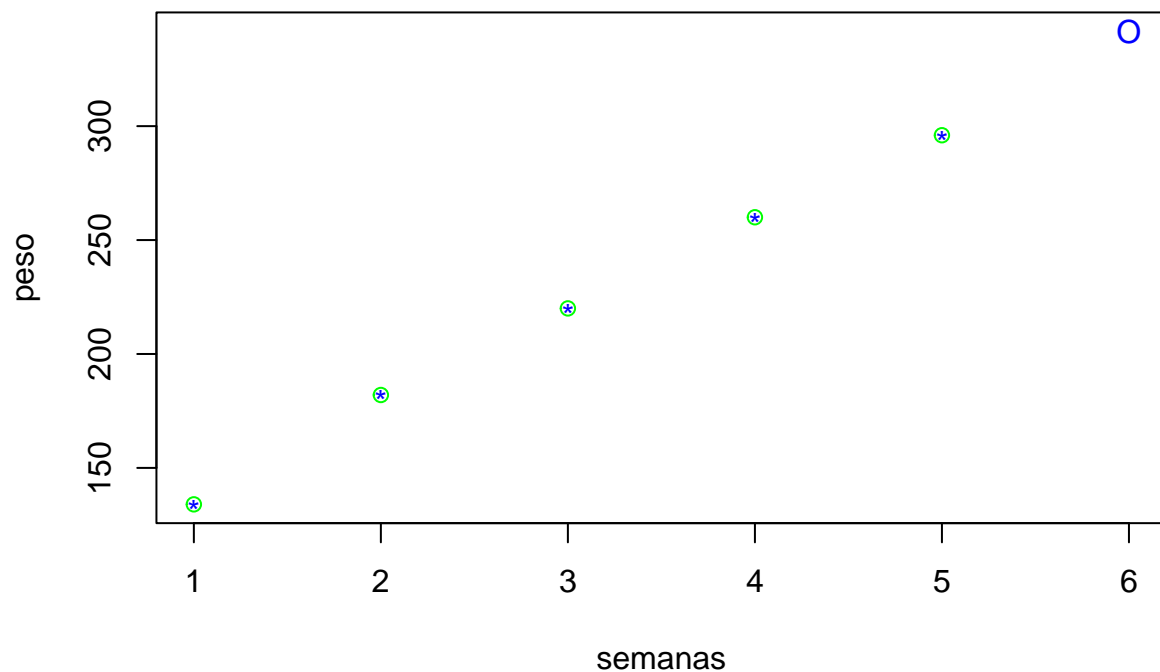
En la siguiente gráfica se muestran los pesos de la rata 10 en las primeras 5 semanas y el peso estimado del día 45 (semana 6).

```
r10_d45_line<-colMeans(rat10[-(1:5000),1:6])

rata10dta<-db[10,2:6]
#rata10dta$x6<-r10_d45_line[6]
rata10dta<-as.numeric(rata10dta)

plot(r10_d45_line, xlab="semanas", ylab="peso", pch = c("*", "*", "*", "*", "*", "0"),
     col="blue")

points(rata10dta, col="green")
```



La distribución predictiva final del peso de la Rata 3 a los 43 días de edad se muestra en la siguiente Gráfica 8.

```
d_rat3<-d[11:15,]
m=10000
sol_drat3=hiergibbs_normlin(d_rat3,m)

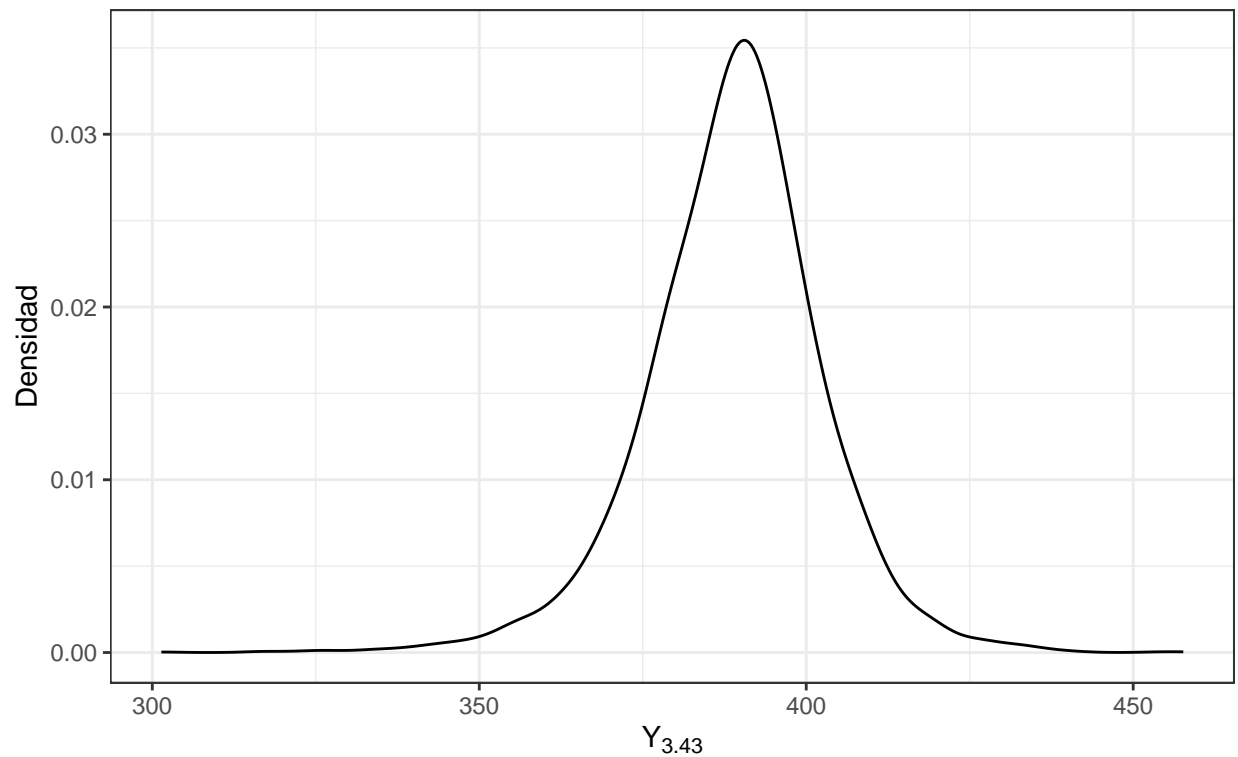
rat3<-sol_drat3$mu
rat3<-as.data.frame(rat3)

prediction2<-sol_drat3$beta[,1]+sol_drat3$beta[,2]*43
prediction2<-as.data.frame(prediction2)

rat3$d43<-prediction2$prediction2

rat3_pred43<-rat3$d43[-(1:5000)]
rat3_pred43<-as.data.frame(rat3_pred43)
p3 <- ggplot(data=rat3_pred43, aes(x=rat3_pred43)) +
  geom_density(adjust=1.5, alpha=.4)+theme_bw() + ylab("Densidad") + ggtitle("Distribución predictiva")
p3
```

Distribución predictiva final del peso de la rata 3 a los 43 días de edad



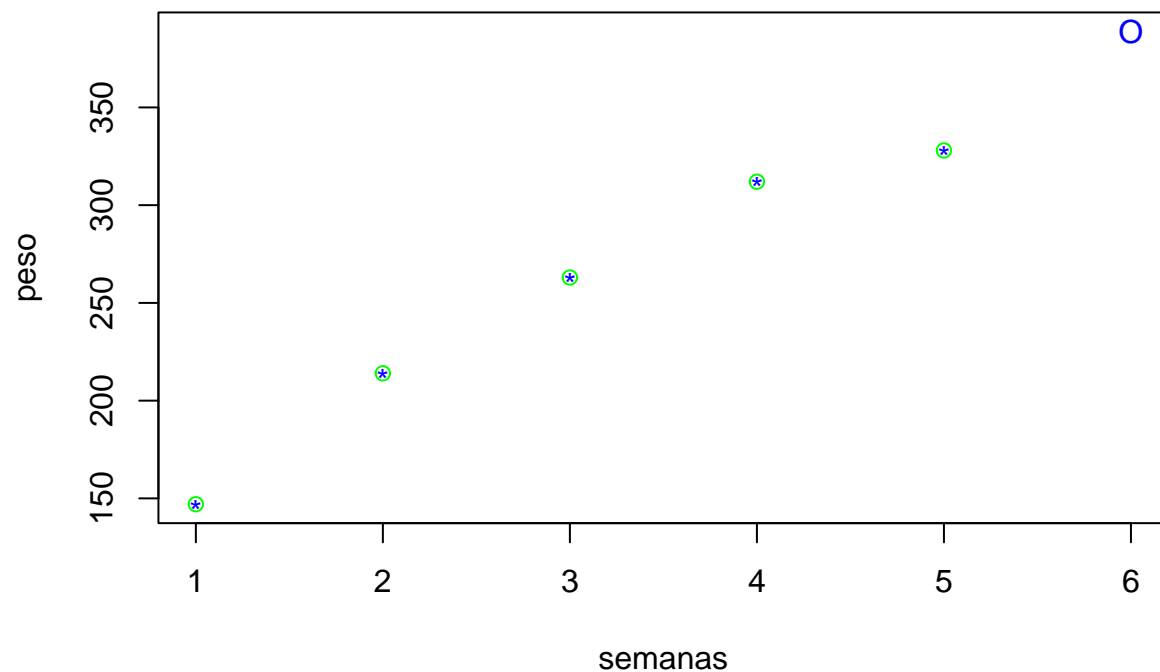
En la siguiente gráfica se muestran los pesos de la rata 3 en las primeras 5 semanas y el peso estimado del día 43 (semana 6).

```
r3_d43_line<-colMeans(rat3[-(1:5000),1:6])

rata3dta<-db[3,2:6]
#rata10dta$x6<-r10_d45_line[6]
rata3dta<-as.numeric(rata3dta)

plot(r3_d43_line, xlab="semanas", ylab="peso", pch = c("*", "*", "*", "*", "*", "0"),
     col="blue")

points(rata3dta, col="green")
```



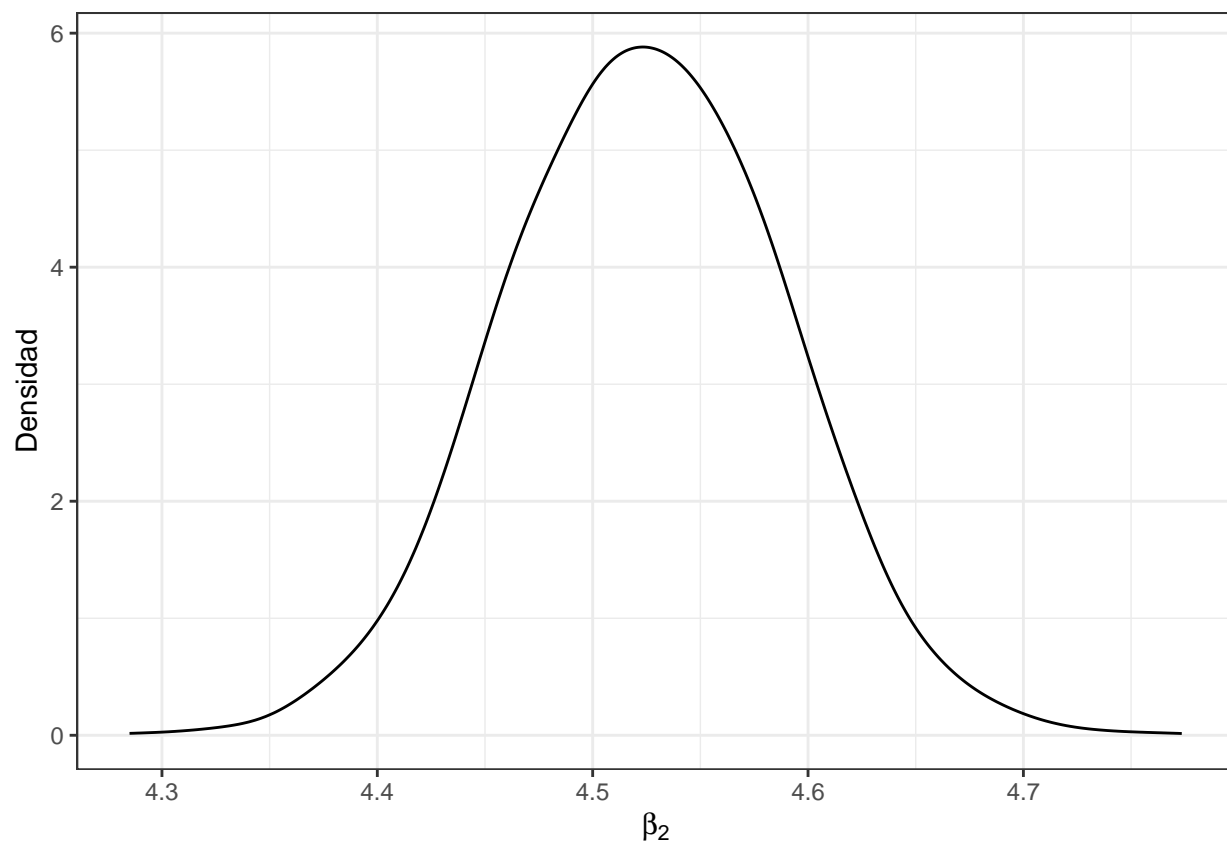
Comparativo con el grupo de tratamiento

Haciendo el mismo análisis para el grupo de tratamiento, tenemos el siguiente resultado con respecto a las diferencias entre las tasas de crecimiento entre el grupo de control y el grupo de tratamiento.

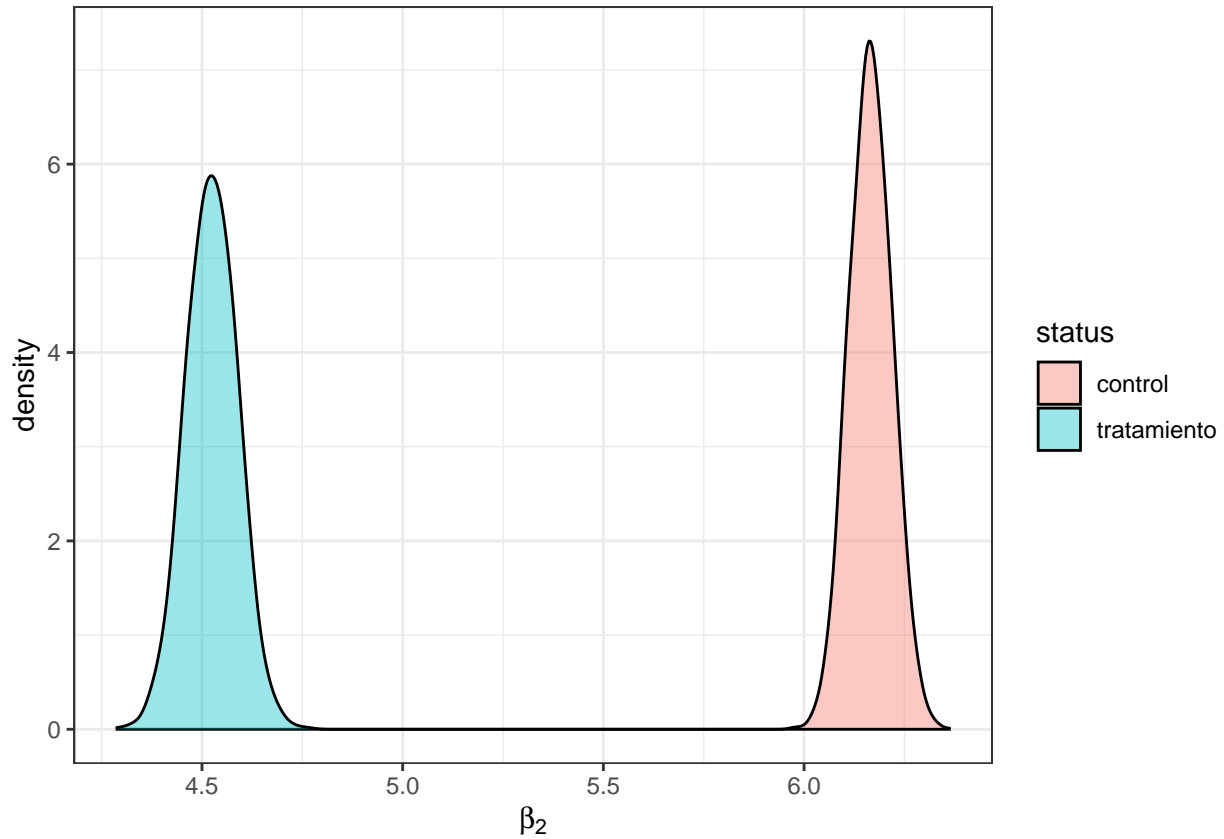
La densidad estimada de la tasa de crecimiento para el grupo de tratamiento se muestra en la Gráfica 8.

```
m=10000
sol2=hiergibbs_normlin(d2,m)

sol_beta2_2<-sol2$beta[-(1:5000),2]
sol_beta2_2<-as.data.frame(sol_beta2_2)
p4 <- ggplot(data=sol_beta2_2, aes(x=sol_beta2_2)) +
  geom_density(adjust=1.5, alpha=.4)+theme_bw()+
  ylab("Densidad") +
  xlab(expression(beta[2]))
p4
```



La comparación entre el grupo de tratamiento y control se muestra en la siguiente Gráfica 9.



Conclusiones

En este documento analizamos teóricamente un modelo lineal normal bayesiano jerárquico con un Gibbs Sampler para calcular densidades posteriores de los parámetros de interés, basado en la propuesta de (Alan E. Gelfand et al., 1990), y lo implementamos computacionalmente con un ejemplo del crecimiento poblacional en grupos de ratas bajo tratamiento y control. El modelo planteado es un modelo jerárquico de tres etapas o capas, considerando algunas “priors” o información a priori, e implementando el método de muestreo de Gibbs. Esto permite un análisis predictivo más flexible tanto para los grupos en su conjunto, como para los individuos (en este caso, ratas) en particular.

Los principales resultados son que la implementación en particular es eficiente, ya que la convergencia se da con pocas iteraciones. Por otra parte, las medias predictivas parecen estar en un nivel adecuado y no se salen de las tendencias y valores observados. Finalmente, se concluye que en general el grupo de tratamiento tuvo una tasa de crecimiento poblacional al rededor del 4.6 %, mucho menor al grupo de control con una tasa al rededor del 6.2 %. Esto muestra que las tasas de crecimiento son distintas entre ambos grupos, y que el tratamiento provoca un efecto significativo en dicha tasa.

Referencias

- Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, & Smith, A. F. M. (1990). *Illustration of bayesian inference in normal data models using gibbs sampling*. Journal of the American Statistical Association. Vol. 85, No. 412, (Dec., 1990), pp. 972-985.
- Albert, J. (2009). *Bayesian computation with r. Second edition*. Springer.
- Albert, J. (May 1, 2019.). *LernBayes package. Function "hiergibbs"*. <https://rdr.io/cran/LearnBayes/src/R/hiergibbs.R>