

Statistiques Bayésiennes

Utilisation du logiciel WinBUGS

- **Partie I** – Prise en main
- **Partie II** – Régression linéaire simple
- **Partie III** – Modèle mixte (Albi)
- **Partie IV** – Diagnostiques de convergence (CODA)
- **Partie V** – Constriction de Doodle (DAG)

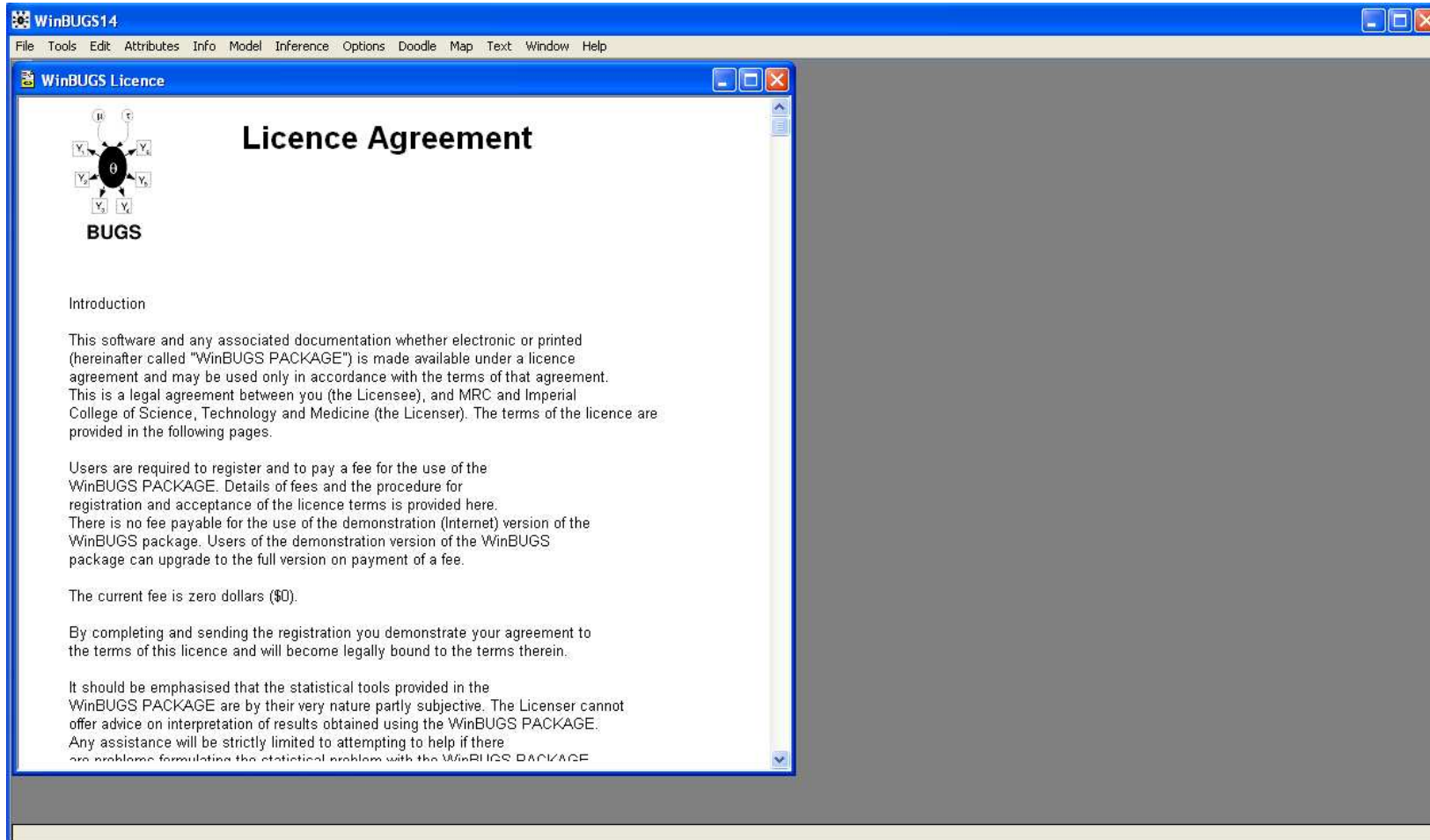
Partie I

Prise en main

WinBUGS

- « **B**ayesian inference **U**sing **G**ibbs **S**ampling »
- The BUGS Project – WinBUGS
<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>
- David Spiegelhalter et al. – Medical Research Council Biostatistics Unit of Cambridge ~ 1996
- Version en développement : OpenBUGS

Prise en main



The screenshot shows the WinBUGS14 software interface. The main window displays the 'WinBUGS Licence' agreement. On the left side of the agreement window, there is a diagram of a Bayesian network with a central node θ and several surrounding nodes labeled Y_1 through Y_6 . Below the diagram is the text 'BUGS'.

Licence Agreement

BUGS

Introduction

This software and any associated documentation whether electronic or printed (hereinafter called "WinBUGS PACKAGE") is made available under a licence agreement and may be used only in accordance with the terms of that agreement. This is a legal agreement between you (the Licensee), and MRC and Imperial College of Science, Technology and Medicine (the Licensor). The terms of the licence are provided in the following pages.

Users are required to register and to pay a fee for the use of the WinBUGS PACKAGE. Details of fees and the procedure for registration and acceptance of the licence terms is provided here. There is no fee payable for the use of the demonstration (Internet) version of the WinBUGS package. Users of the demonstration version of the WinBUGS package can upgrade to the full version on payment of a fee.

The current fee is zero dollars (\$0).

By completing and sending the registration you demonstrate your agreement to the terms of this licence and will become legally bound to the terms therein.

It should be emphasised that the statistical tools provided in the WinBUGS PACKAGE are by their very nature partly subjective. The Licensor cannot offer advice on interpretation of results obtained using the WinBUGS PACKAGE. Any assistance will be strictly limited to attempting to help if there are problems formulating the statistical problem with the WinBUGS PACKAGE.

Prise en main (MENU)

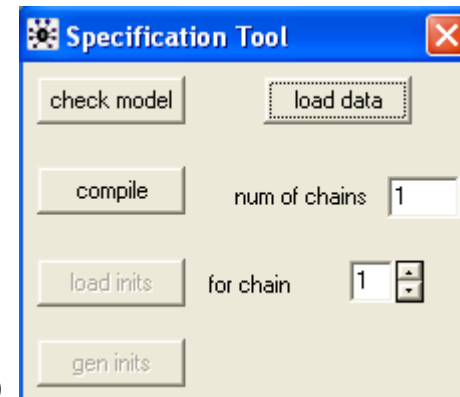
- **File : Ouvrir, sauver, fermer ...**
- **Tools** : Insertion formules, encodage, décodage...
- **Edit** : Copier/Coller sélectionner
- **Attributes** : Police du texte
- **Info** : Log et information sur les nœuds
- **Model : Définition du modèle**
- **Inference : Estimation, convergence ... sur le modèle**
- **Options**: Affichage et update
- **Doodle** : Représentation graphique du modèle (Directed Acyclic Graphs)
- **Map** : Cartographie
- **Text** : Mise en forme du texte (recherche, paragraphes...)
- **Window** : Gestionnaire des fenêtres.
- **Help: Manuels, tutoriaux et exemples.**

Prise en main (FENETRES)

- **Log (Info -> Open log) :**
 - Information de compilation
 - Sortie des résultats dans la LogOptions->Output options -> cocher log



- Fenêtre de script (File -> New)
- Fenêtre de spécification du modèle
- Fenêtres pop-up de résultats
- Bandeau inférieur (déroulement du programme)



Prise en main (VIDEO EXAMPLE)



Welcome to **WinBUGS** - the movie!

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/winbugsthemovie.html>

Prise en main (PROGRAMMATION)

- Langage semblable à **R**
- Attention aux **paramétrisation** -> Voir **Distributions** dans Help
 - Exemple de la loi normale $N(\mu, \tau^2) \sim N(\mu, 1/\sigma^2)$
- Choix des **a priori** :
 - WinBUGS n'autorise pas d'a priori impropres.
 - Le manuel propose des a priori non informatifs.
 - Paramètres de position μ : A priori $N(0, 0.0001)$**
 - Paramètres d'échelle τ^2 : A priori $G(0.0001, 0.0001)$**

Un exemple pour se convaincre

- Simulation puis estimation des paramètres d'une loi normale:

- Simulation R : Loi Normale $N(102,10)$

```
a<-rnorm(150, mean=102, sd=10)
setwd("P:/")
write.table(t(a),file="dataNORMALE.txt",
  sep=',',row.names=FALSE,col.names=FALSE)
```

- Estimation Bayésienne des paramètres : μ et σ^2

Cas Pratique Spécification

Modèle

->

Données

->

Valeurs initiales

```
#MODEL
model{
  # Vraisemblance
  for (i in 1:N){
    Qi[i]~dnorm(mu,tausq)
  }
  # A priori
  mu~dnorm(0,0.0001)
  tausq~dgamma(0.0001,0.0001)
  # VariableInteret
  sigmasq<-1/tausq
}
```

```
#DATA
list(simul=c(106.017394973321,
118.203182695254,...,
111.889260634063,...,
84.8659649363283),N=75)
```



Data_gauss.txt

```
#INITS CHAINE 1
list(mu=0,tausq=1)
#INITS CHAINE 2
list(mu=200,
      tausq=1000)
```

Compilation (nombre de chaînes) – 2 stratégies :

- 1 chaîne très longue / avec éventuellement un échantillonnage espacé (thin)
- Plusieurs chaînes plus courtes

Cas Pratique Méthodologie - Update and Sample

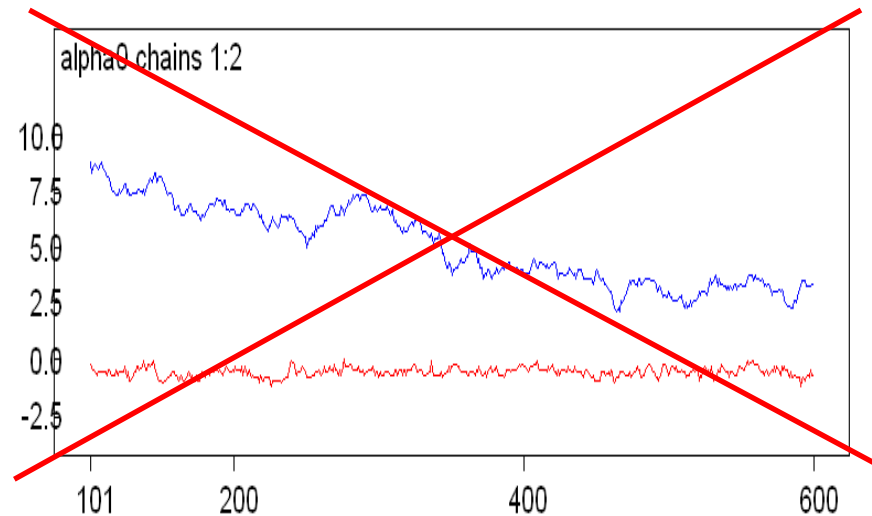
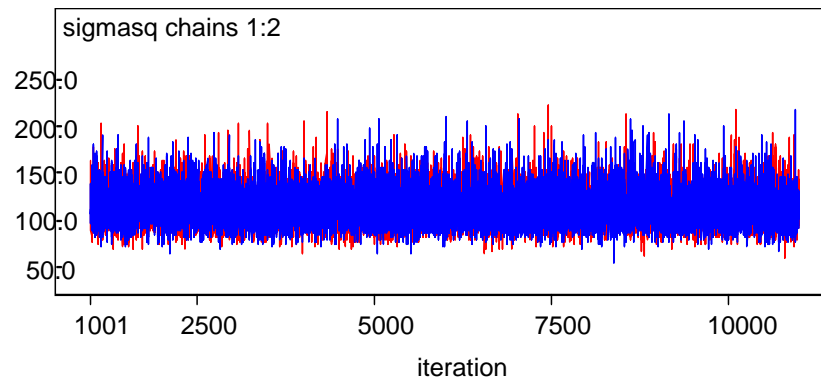
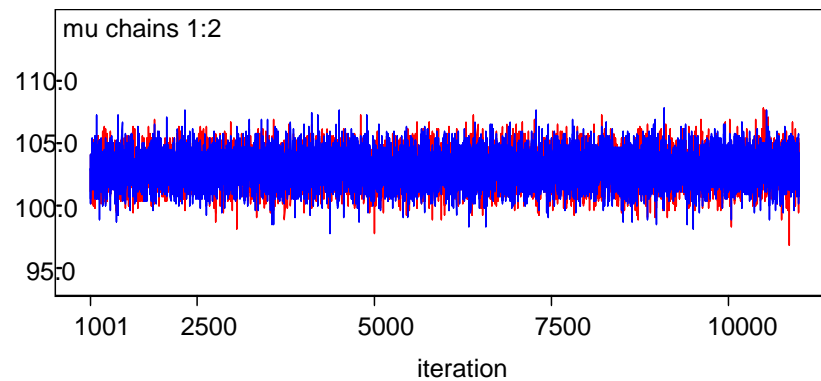
- **Update (1) Burn-in phase (ici 1000):**
 - Simulations de la phase de chauffe
 - Les valeurs prises par les paramètres ne sont pas sauvegardées.
 - La longueur requise dépend du problème.

- **Inférence**
 - **Sample** : définition des nœuds (ici μ , σ^2)
 - **Set/Clear** : Ajouter/Supprimer le nœud
 - **DIC** : Set

- **Update (2) Echantillonnage d'intérêt (ici 10 000)**

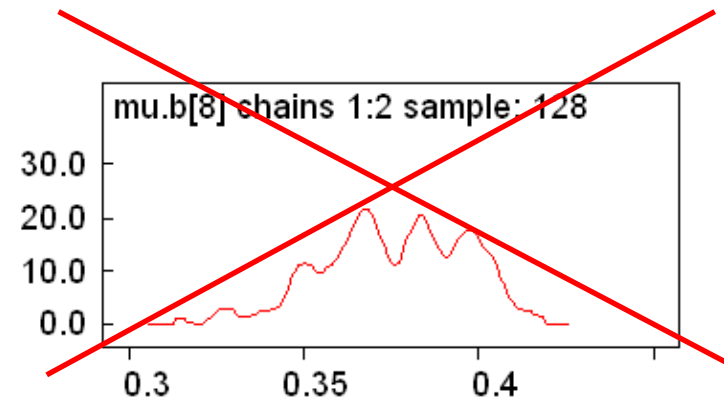
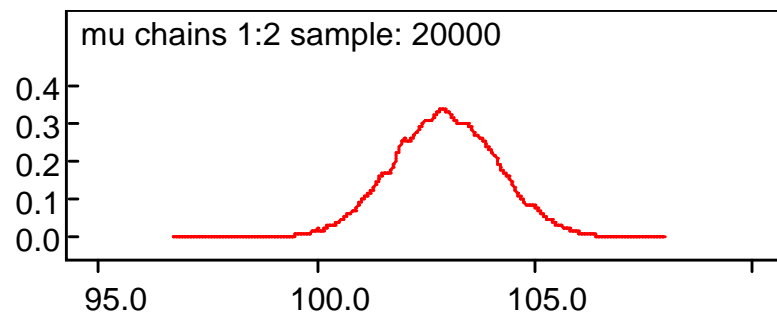
Cas Pratique Résultats a posteriori

- **Sample** : sélection du nœud ou *
- **History/Trace** : Evaluation de la convergence (mélangeance)

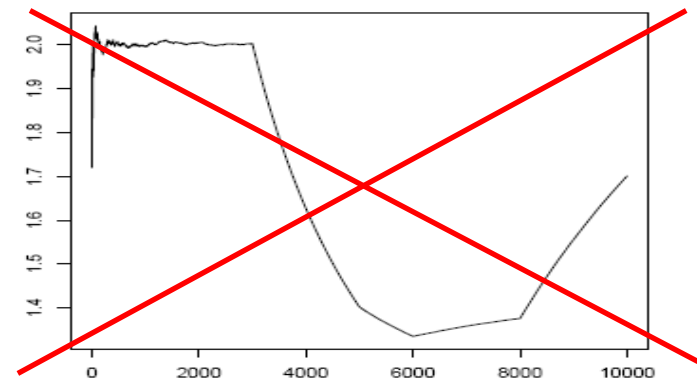
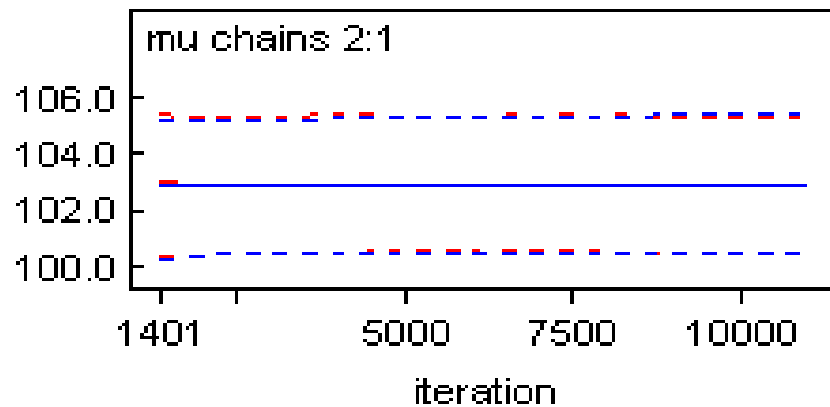


Cas Pratique Résultats a posteriori

■ Density : Densité a posteriori



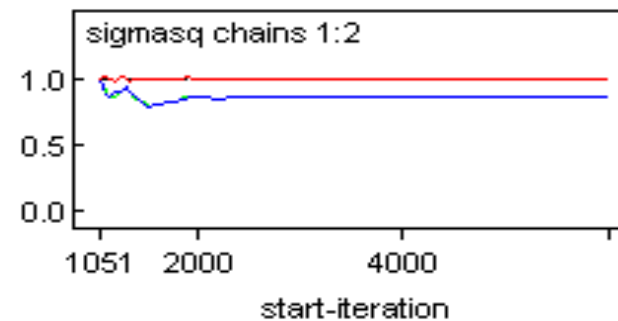
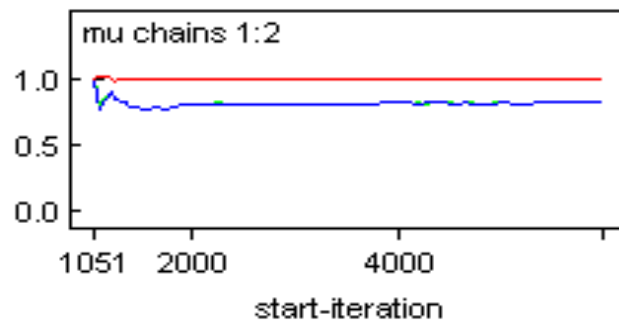
■ Quantiles :



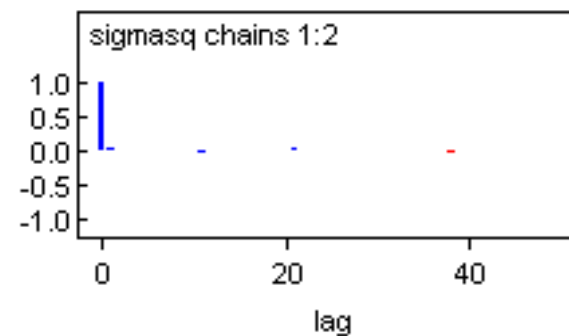
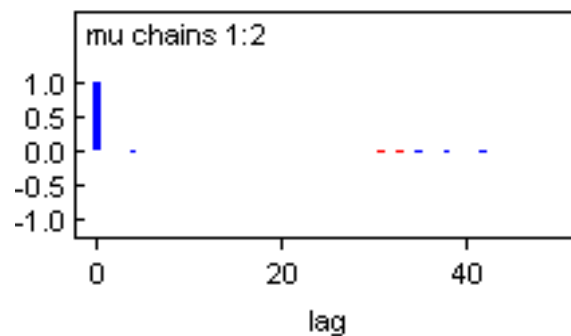
Cas Pratique Résultats a posteriori

- **Bgr diag** : Calcule la stat. de Gelman-Rubin.

Convergence si $R < 1.05$ et stabilité autour de 1. B et W doivent aussi s'être stabilisés



- **Auto corr** : Fonction d'auto corrélation de la variable avec un lag-50.
Solution si auto corrélé : augmenter l'épaisseur d'échantillonnage (thin)



Cas Pratique Résultats a posteriori

- **Stats** : Statistiques descriptives de la loi a posteriori

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mu	102.9	1.232	0.008973	100.5	102.9	105.4	1001	20000
sigmasq	114.5	19.41	0.1295	82.79	112.3	158.0	1001	20000

- **Coda** : Valeurs des chaines en sortie texte pour analyser la convergence avec R (package Coda – voir lundi 12/12/11)

Cas Pratique Résultats a posteriori

- **DIC** : Deviance Information Criterion

It is important to note that DIC assumes the posterior mean to be a good estimate of the stochastic parameters. If this is not so, say because of extreme skewness or even bimodality, then DIC may not be appropriate.

Dbar = post.mean of $-2\log L$; **Dhat** = $-2\log L$ at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
simul	567.309	565.290	2.018	569.327
total	567.309	565.290	2.018	569.327

Dbar: Moyenne a posteriori de la deviance définie comme $-2 * \log(\text{likelihood})$: 'likelihood' est calculée par $p(y | \theta)$.

Dhat: Estimation ponctuelle de la déviance. Obtenu en substituant θ par sa moyenne a posteriori.

$$\text{Dhat} = -2 * \log(p(y | \theta.\text{bar})).$$

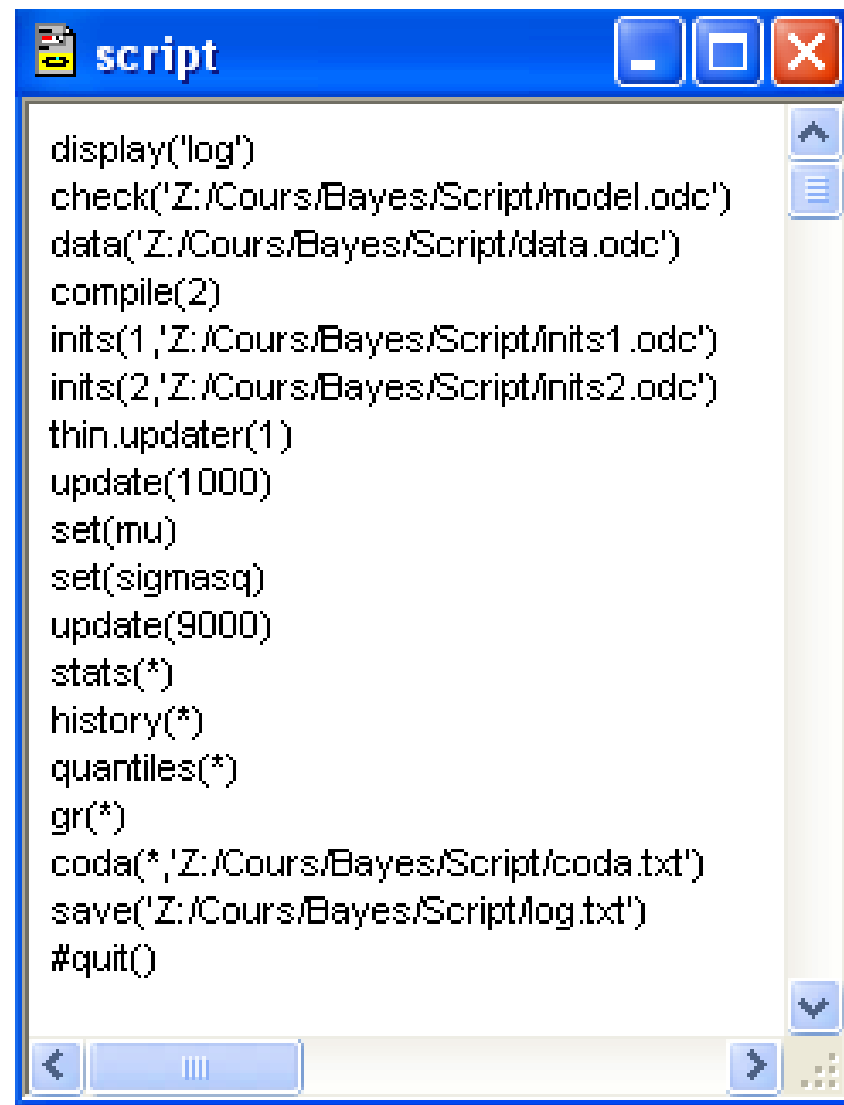
pD: Nombre effectif de paramètres. $pD = \text{Dbar} - \text{Dhat}$

$$\text{DIC} = \text{Dbar} + pD = \text{Dhat} + 2 * pD$$

Plus le **DIC est petit** plus le modèle est le meilleur prédicteur pour un jeux de données de la même structure.

Fonctionnalité avancées

- Scripts



```
display('log')
check('Z:/Cours/Bayes/Script/model.odc')
data('Z:/Cours/Bayes/Script/data.odc')
compile(2)
inits(1,'Z:/Cours/Bayes/Script/inits1.odc')
inits(2,'Z:/Cours/Bayes/Script/inits2.odc')
thin.updater(1)
update(1000)
set(mu)
set(sigmasq)
update(9000)
stats(*)
history(*)
quantiles(*)
gr(*)
coda(*,'Z:/Cours/Bayes/Script/coda.txt')
save('Z:/Cours/Bayes/Script/log.txt')
#quit()
```

Winbugs et autres logiciels

Des communications depuis d'autres logiciels sont possibles, soit :

- Pour analyser les résultats
- Pour lancer Winbugs depuis une autre interface
- Pour se substituer à Winbugs

■ **R**

- Coda (Plummer et al.) – Diagnostic de convergence
- Brugs, JAGS (interfacé avec C), R2WinBugs ...

■ **SAS**

- Macro SAS (Hayat et al., Zhang et al. ...)

Partie II
Régression linéaire
simple

Modèle de Régression Linéaire

- **Objectif** : Expliquer le poids de cochon d'élevage en fonction de leur âge et du nombre de ration de nourriture journalière.

Cochon	1	2	3	4	5	6	7	8	9	10
Poids en kg (P)	85	112,5	94,5	123,5	107	85	101	140	110,5	99,5
Ration/Jour (R)	4	3	3	2	4	3	2	2	4	3
Age en mois (A)	143	191	160	157	191	141	185	210	149	169

$$p_i = \beta_1 a_i + \beta_2 r_i + \varepsilon_i; i = 1, \dots, 10$$

$$\varepsilon_i \sim N(0, \tau)$$

$$p_i \sim N(\beta_1 a_i + \beta_2 r_i, \tau)$$

A faire en WinBUGS et R

- **Approche Bayésienne (WinBUGS)**
 - Ecrire et « Check model » (spécification et a priori)
 - Charger les données
 - Compiler et initialiser le modèle :
 - Nombre de chaînes MCMC
 - Valeurs initiales
 - Simuler la phase de chauffe
 - Entrer les paramètres à échantillonner
 - Lancer l'échantillonnage pour générer la loi a posteriori
 - Vérifier la convergence
 - Afficher les résultats

- **Approche Fréquentiste (R)**
- **Comparaison des deux approches**

Code WinBUGS et R

■ Approche Bayésienne (WinBUGS)

```
model{
  for (i in 1:N){
    p[i]~dnorm(mu[i],tausq)
    mu[i] <-
      beta1*a[i]+beta2*r[i]
  }
  #A priori
  beta1~dnorm(0,0.0001)
  beta2~dnorm(0,0.0001)
  tausq~dgamma(0.001,0.001)
  # Variable d'intérêt
  sigmasq<-1/tausq
}
```

#DATA

```
list(N=10,p=c(85,112.5,94.5,
123.5,107,85,101,140,110.5,
99.5),r=c(4,3,3,2,4,3,2,2,4,3),
a=c(143,191,160,157,191,141,
185,210,149,169))
```

#INITS1

```
list(beta1=0,beta2=0,
      tausq=1)
```

#INITS1

```
list(beta1=4,beta2=4,
      tausq=4)
```

■ Approche Fréquentiste (R)

```
lmall<-lm(P~R+A-1)
summary(lmall)
```

```
P<-c(85,112.5,94.5,123.5,107,85,101,140,110.5,99.5)
R<-c(4,3,3,2,4,3,2,2,4,3)
A<-c(143,191,160,157,191,141,185,210,149,169)
```

Résultats

- Approche Bayésienne (WinBUGS)

convergence OK.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	0.6338	0.09139	6.449E-4	0.4518	0.634	0.8146	1001	20000
beta2	-0.7179	5.039	0.03649	-10.72	-0.7676	9.41	1001	20000
sigma	15.28	4.446	0.04057	9.36	14.41	26.1	1001	20000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
p	82.155	78.885	3.271	85.426
total	82.155	78.885	3.271	85.426

- Approche Fréquentiste (R)

Call:

```
lm(formula = P ~ R + A - 1)
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
R -0.70800    4.38468  -0.161    0.876
A  0.63388    0.07942   7.982 4.44e-05 ***
```

Residual standard error: 13.81 on 8 degrees of freedom

Multiple R-Squared: 0.9867, Adjusted R-squared: 0.9834

F-statistic: 296.7 on 2 and 8 DF, p-value: 3.132e-08

Nouveau modèle de régression

$$p_i = \beta_1 a_i + \varepsilon_i; i = 1, \dots, 10$$

$$\varepsilon_i \sim N(0, \tau)$$

$$p_i \sim N(\beta_1 a_i, \tau)$$

- Approche Bayésienne (WinBUGS) – Convergence OK

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	0.6216	0.02729	1.903E-4	0.5662	0.6216	0.6763	1	20000
sigma	218.5	137.4	1.021	80.31	184.2	561.1	1	20000

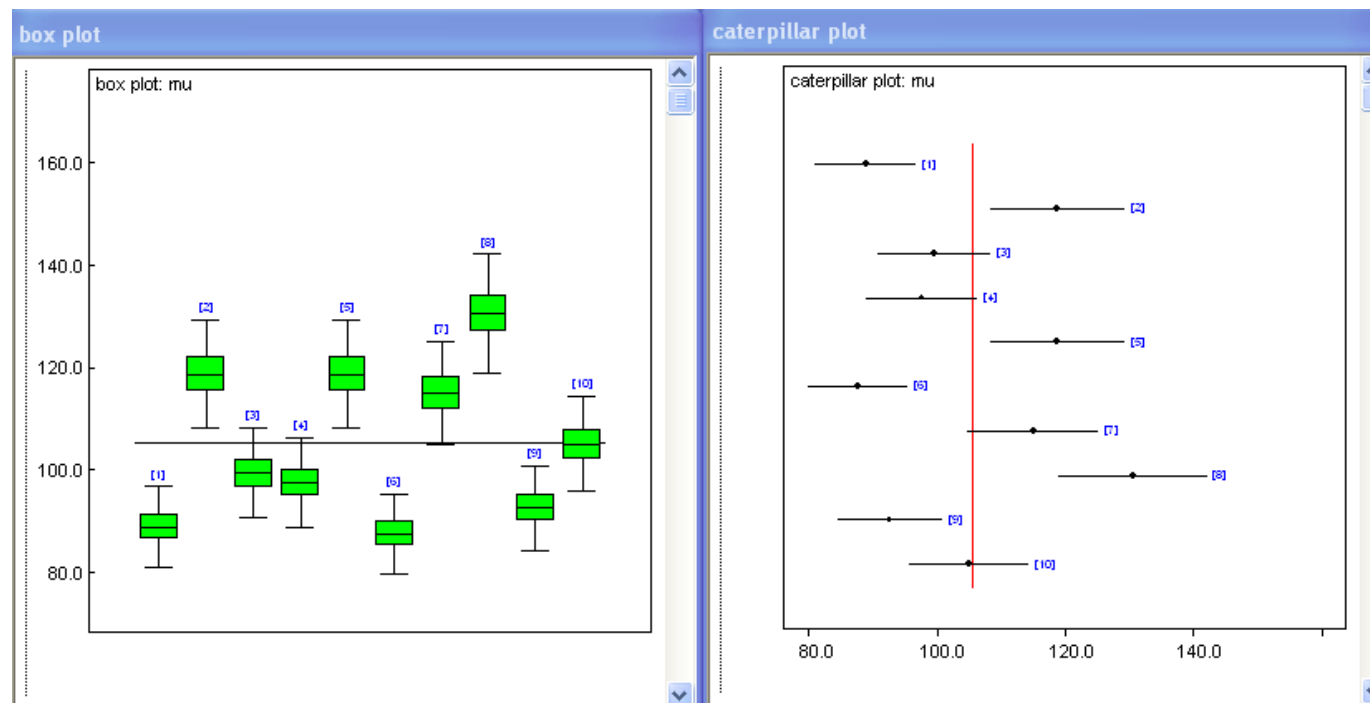
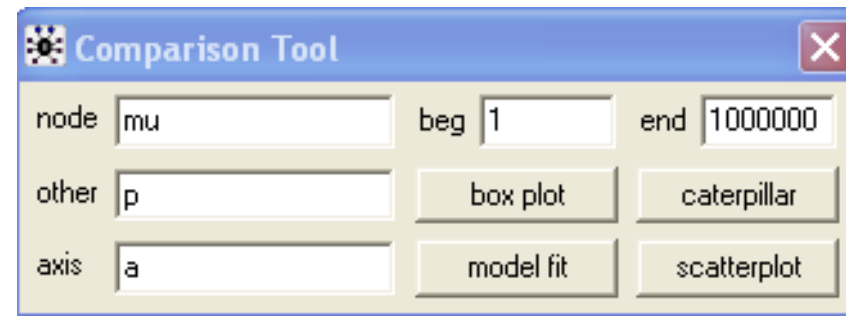
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
p	80.894	78.739	2.155	83.049
total	80.894	78.739	2.155	83.049

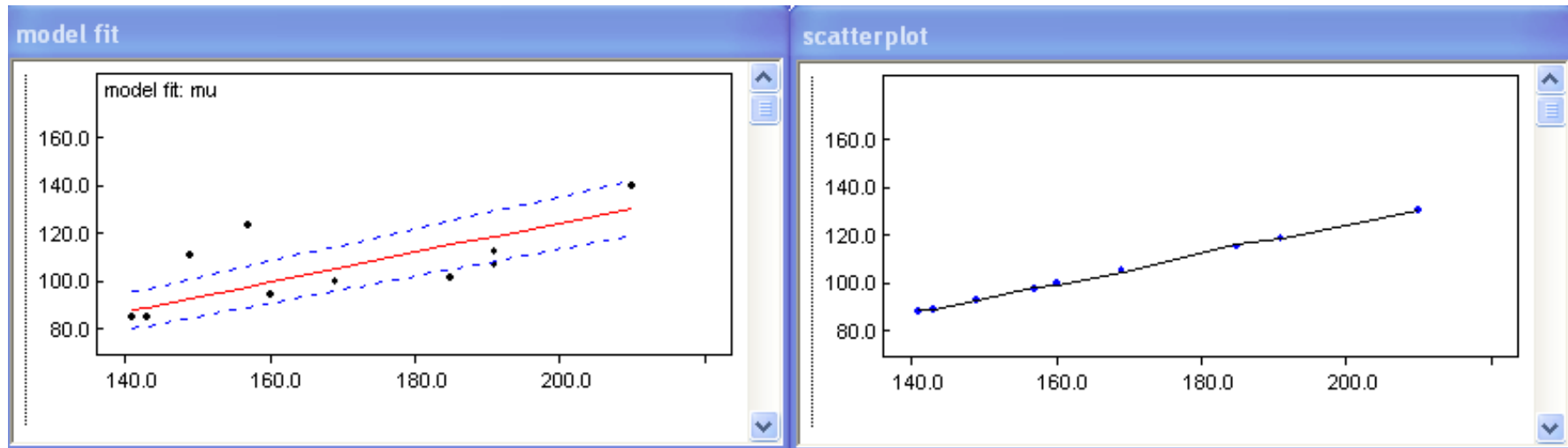
- Approche Fréquentiste (R)

```
Call:
lm(formula = P ~ A - 1)
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
A    0.6217     0.0241   25.79 9.54e-10 ***
Residual standard error: 13.04 on 9 degrees of freedom
Multiple R-Squared:  0.9867,    Adjusted R-squared:  0.9852
F-statistic: 665.3 on 1 and 9 DF,  p-value: 9.536e-10
```

Courbes de prédiction



Courbes de prédiction



- Il est aussi possible de récupérer :
 - Les résidus
 - Les empirical Bayes pour les variables avec des effets aléatoires
 - MSE ...

Partie III

Modèle mixte (Albi)

L'essai thérapeutique ALBI-ANRS 070

L'essai thérapeutique ALBI - ANRS 070 était un essai comparant l'efficacité et la tolérance de trois stratégies antirétrovirales à 6 mois chez des patients infectés par le VIH-1 et naïfs de tout traitement antirétroviral (Molina, Chêne et al. 1999). Les autres critères d'inclusion étaient 1) un taux de CD4 supérieur ou égal à 200 lymphocytes/mm³ et 2) une charge virale comprise entre 10 000 et 100 000 copies/ml.

Les patients étaient randomisés dans l'un des trois bras de traitement suivant :

- bras 'AZT+3TC' : les patients étaient traités pendant 24 semaines par une bithérapie d'analogues nucléosidiques, zidovudine (AZT) et lamivudine (3TC),
- bras 'd4T+ddI' : les patients étaient traités par une bithérapie de stavudine (d4T) et didanosine (ddI) pendant 24 semaines,
- bras 'alternance' : les patients étaient traités 12 semaines par une bithérapie associant stavudine et didanosine puis 12 semaines par zidovudine et lamivudine.

Figure 1. Evolution des lymphocytes T CD4+ selon le groupe de traitement.

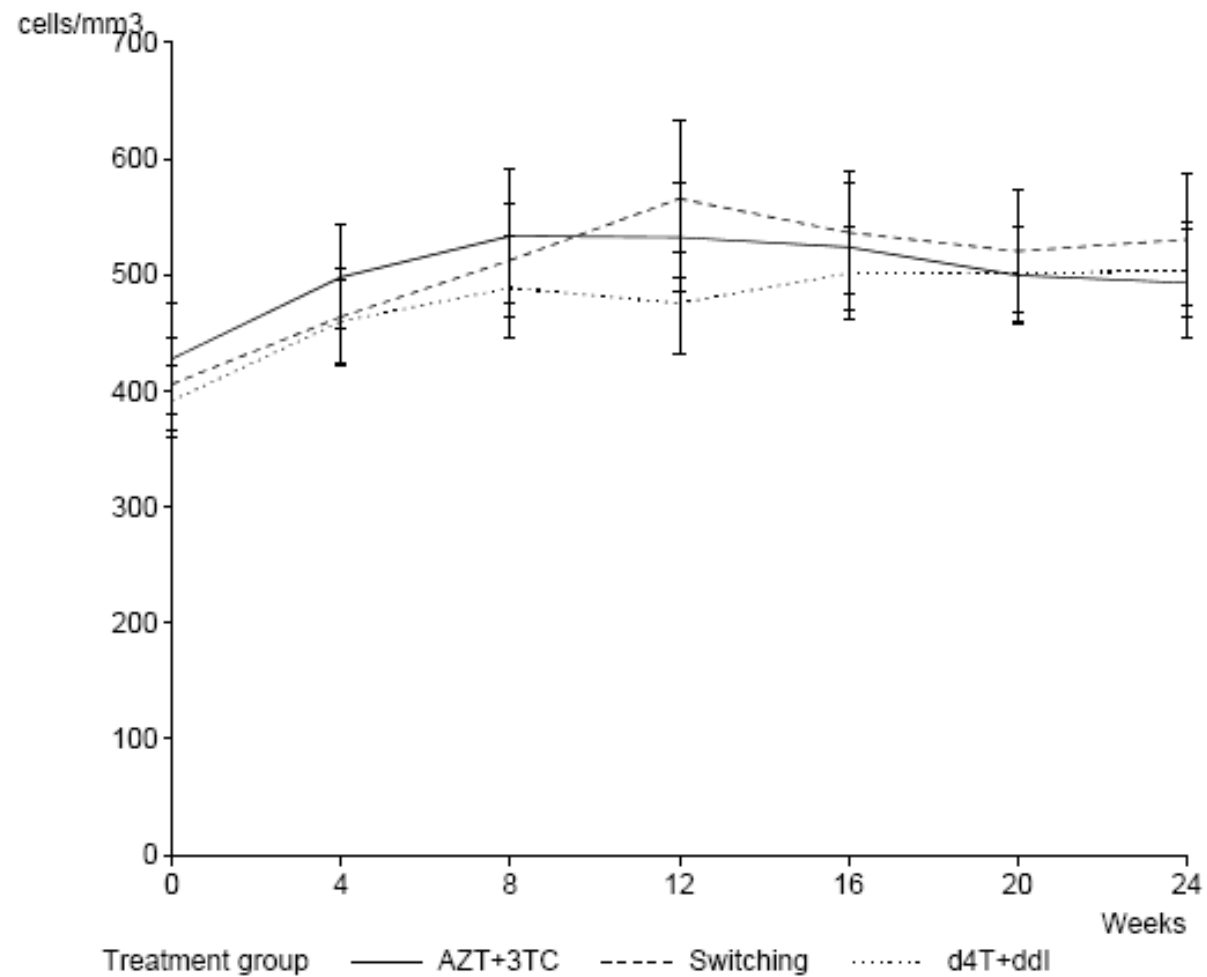
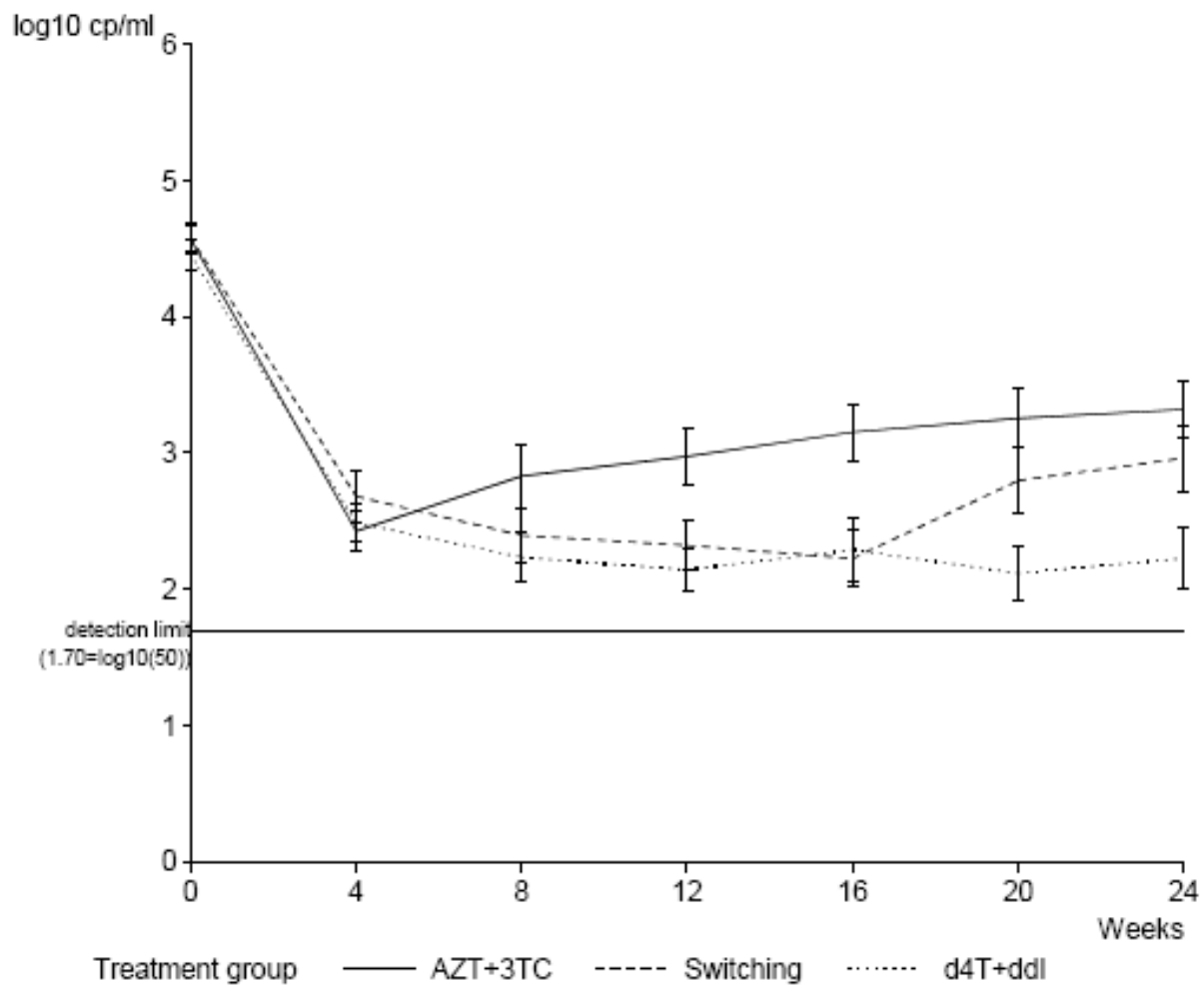


Figure 2. Evolution de la charge virale plasmatique selon le groupe de traitement.




ALBI

	NUM_PAT	VIS	TD	T	RAN_GRP	CD4	CD4t	LCV_50
1	15	0	0	0	3	258	4.00778972	3.914
2	15	1	4	0.28	3	390	4.44391918	2.431
3	15	2	8	0.56	3	322	4.23607819	2.061
4	15	3	12	0.85	3	482	4.68556276	1.699
5	15	4	16	1.14	3	443	4.58776255	2.253
6	15	5	20	1.41	3	606	4.96155895	1.845
7	15	6	24	1.63	3	437	4.57214883	1.978
8	19	0	0	0	2	392	4.44960559	4.504
9	19	1	4	0.29	2	318	4.22286094	4.068
10	19	2	8	0.56	2	508	4.74751044	2.924
11	19	3	12	0.84	2	539	4.81833722	2.248
12	19	4	16	1.12	2	573	4.89258811	1.748
13	19	5	20	1.38	2	385	4.42960685	4.117
14	19	6	24	1.73	2	303	4.17215714	2.96
15	24	0	0	0	1	390	4.44391918	4.639
16	24	1	4	0.31	1	378	4.40933352	2.26
17	24	2	8	0.59	1	468	4.65116197	1.699
18	24	3	12	0.91	1	480	4.68069464	2.982
19	24	4	16	1.19	1	462	4.63618219	3.159
20	24	5	20	1.5	1	438	4.57476224	2.82
21	24	6	24	1.82	1	479	4.67825487	2.76
22	33	0	0	0	3	445	4.59293186	4.821
23	33	1	4	0.28	3	400	4.47213595	2.097
24	33	2	8	0.56	3	720	5.18004013	1.903
25	33	4	16	1.19	3	738	5.21211621	1.875
26	33	5	20	1.53	3	632	5.01394158	1.699
27	33	6	24	1.87	3	621	4.99198073	1.699

Data_albi.xls et albi_winbugs.txt

ALBI : variables disponibles

Variable	Description
CD4	Nombre de lymphocytes T CD4+ en cellules / mm ³
CD4t	Racine quatrième du nombre de lymphocytes T CD4+
LCV_50	Charge virale plasmatique
T	Délai entre la visite et l'inclusion en jours divisé par 100
VIS	Numéro de la visite
TD	Délai entre la visite et l'inclusion en semaines
NUM_PAT	Identifiant patient
RAN_GRP	Groupe de traitement (1, 2 ou 3)

- 
- 1 : d4t+ddI
 - 2 : Switch
 - 3 : AZT+3TC

Modèle à intercept aléatoire

- Temps discret (0 jours, 30 jours, 60 jours, ... ,180 jours)

$$y_{ij} = \beta_0 + \beta_1 t_j + \gamma_{0i} + \varepsilon_{ij}$$

$$\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$$

$$(\gamma_{01} \dots \gamma_{0n}) \perp (\varepsilon_{11} \dots \varepsilon_{1n})$$

$$y_{ij} = \sqrt[4]{CD4_{ij}}$$

- Temps continu

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{0i} + \varepsilon_{ij}$$

$$\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$$

$$(\gamma_{01} \dots \gamma_{0n}) \perp (\varepsilon_{11} \dots \varepsilon_{1n})$$

$$y_{ij} = \sqrt[4]{CD4_{ij}}$$

Modèle à intercept aléatoire (temps discret)

- Temps discret

```
# Definition du modele
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      CD4t[i, j] ~ dnorm(mu[i, j], tau.eps);
      mu[i, j] <- beta0 + beta1*tps[j]+gamma0[i]
    }
    gamma0[i] ~ dnorm(0.0, tau.gamma0)
  }
}

# priors:
beta0 ~ dnorm(0.0, 1.0E-4)
beta1 ~ dnorm(0.0, 1.0E-4)
tau.eps ~ dgamma(1.0E-3, 1.0E-3)
var.eps <- 1.0/tau.eps
tau.gamma0 ~ dgamma(1.0E-3, 1.0E-3)
var.gamma0 <- 1.0 /tau.gamma0
}
```

Modèle à intercept aléatoire (temps discret)

- Temps discret

```
list(  
  N=149,  
  T=7,  
  CD4t = structure(.Data =  
    c(4.0077897156,4.4439191777,....,NA,....,4.8205705137),.Dim = c(149,7)),  
  tps = c(0,0.30,0.60,0.90,1.20,1.50,1.80))
```

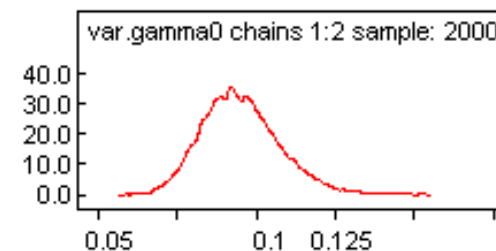
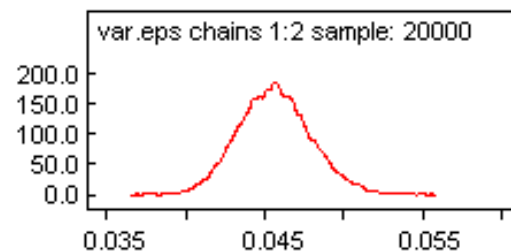
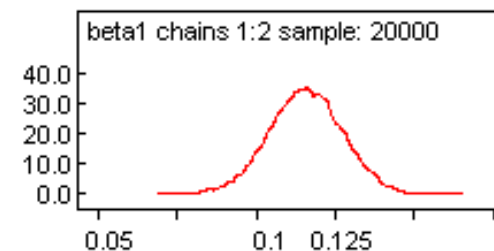
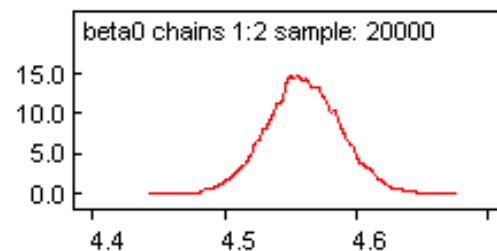
Modèle à intercept aléatoire (temps discret)

■ Temps discret

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.56	0.02826	9.022E-4	4.504	4.56	4.616	1001	20000
beta1	0.1157	0.01149	8.807E-5	0.09296	0.1157	0.138	1001	20000
var.eps	0.04564	0.00228	1.696E-5	0.04134	0.04559	0.05033	1001	20000
var.gamma0	0.09512	0.01215	1.02E-4	0.07386	0.09415	0.1215	1001	20000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
CD4t	-237.910	-378.000	140.090	-97.820
total	-237.910	-378.000	140.090	-97.820



Modèle à intercept aléatoire SAS - proc MIXED (temps discret)

■ Temps discret

node	mean	sd	\bar{D} = post.mean of $-2\log L$; D_{hat} = $-2\log L$ at post.mean of stochastic nodes				
beta0	4.56	0.02826					
beta1	0.1157	0.01149					
var.eps	0.04564	0.00228	CD4t	Dbar	Dhat	pD	DIC
var.gamma0	0.09512	0.01215	total	-237.910	-378.000	140.090	-97.820

Parm cov	Sujet	Estimation
Intercept	NUM_PAT	0.09297
Residual		0.04546

Statistiques d'ajustement

-2 Log Likelihood	149.9
AIC (smaller is better)	157.9
AICC (smaller is better)	158.0
BIC (smaller is better)	169.9

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	4.5601	0.02785	148	163.71	<.0001
TVIS	0.1156	0.01149	800	10.06	<.0001

Modèle à intercept aléatoire (temps continu)

- Temps continu

```
# Definition du modele
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      CD4t[i, j] ~ dnorm(mu[i, j], tau.eps);
      mu[i, j] <- beta0 + beta1*tps[i, j] + gamma0[i]
    }
    gamma0[i] ~ dnorm(0.0, tau.gamma0)
  }
}

# priors:
beta0 ~ dnorm(0.0, 1.0E-4)
beta1 ~ dnorm(0.0, 1.0E-4)
tau.eps ~ dgamma(1.0E-3, 1.0E-3)
var.eps <- 1.0/tau.eps
tau.gamma0 ~ dgamma(1.0E-3, 1.0E-3)
var.gamma0 <- 1.0 /tau.gamma0
}
```

Modèle à intercept aléatoire (temps continu)

- DATAS (temps continu)

```
list(  
  N=149,  
  T=7,  
  CD4t = structure(.Data =  
    c(4.0077897156,4.4439191777,...,NA,...,4.8205705137),.Dim = c(149,7)),  
  tps = structure(.Data = c(0,0.28,0.56,0.85,...,1.81),.Dim = c(149,7)))
```

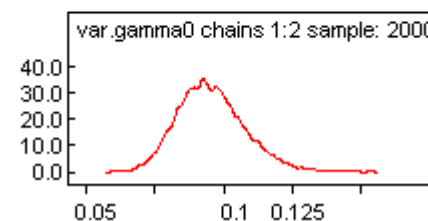
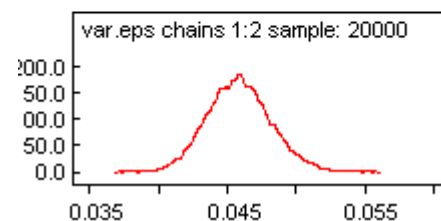
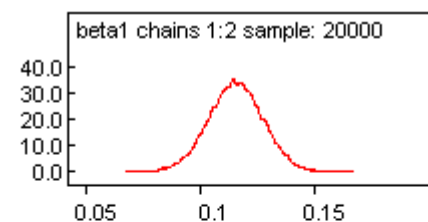
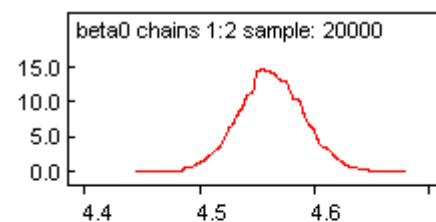

Modèle à intercept aléatoire (temps continu)

■ Temps continu

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.563	0.02832	9.046E-4	4.507	4.562	4.619	1001	20000
beta1	0.1153	0.01173	9.127E-5	0.092	0.1153	0.1381	1001	20000
var.eps	0.04586	0.002291	1.704E-5	0.04154	0.04581	0.05057	1001	20000
var.gamma0	0.09556	0.01221	1.024E-4	0.07422	0.09459	0.1221	1001	20000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
CD4t	-233.354	-373.441	140.086	-93.268
total	-233.354	-373.441	140.086	-93.268



Modèle à intercept aléatoire SAS - proc MIXED (temps continu)

- Temps continu

node	mean	sd	Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes				
				Dbar	Dhat	pD	DIC
beta0	4.563	0.02832					
beta1	0.1153	0.01173					
var.eps	0.04586	0.002291	CD4t	-233.354	-373.441	140.086	-93.268
var.gamma0	0.09556	0.01221	total	-233.354	-373.441	140.086	-93.268

Covariance Parameter Estimates

Parm cov	Sujet	Estimation
Intercept	NUM_PAT	0.09340
Residual		0.04568

Statistiques d'ajustement

-2 Log Likelihood	154.5
AIC (smaller is better)	162.5
AICC (smaller is better)	162.5
BIC (smaller is better)	174.5

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	4.5626	0.02791	148	163.47	<.0001
T	0.1152	0.01174	800	9.81	<.0001

Modèle à intercept et pente aléatoires

$$\left\{ \begin{array}{l} y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij} \\ \gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\gamma_0}^2 & 0 \\ 0 & \sigma_{\gamma_1}^2 \end{bmatrix} \right) \\ \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2) \\ (\gamma_1 \dots \gamma_n) \perp (\varepsilon_1 \dots \varepsilon_n) \\ y_{ij} = \sqrt[4]{CD4_{ij}} \end{array} \right.$$

Modèle à intercept et pente aléatoires

```
# Definition du modele
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      CD4t[i, j] ~ dnorm(mu[i,j], tau.eps);
      mu[i,j] <- beta0 + beta1*tps[i,j]+ gamma0[i] + gamma1[i]* tps[i,j]
    }
    gamma0[i] ~ dnorm(0.0,tau.gamma0)
    gamma1[i] ~ dnorm(0.0,tau.gamma1)
  }
}

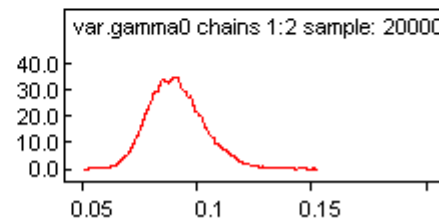
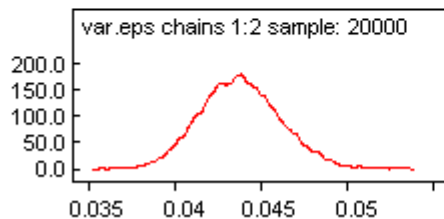
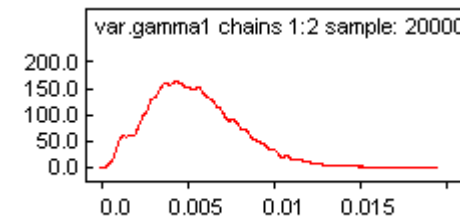
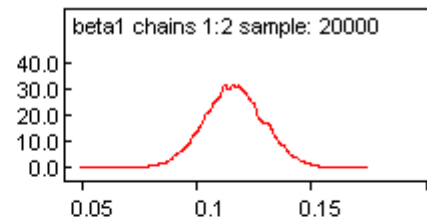
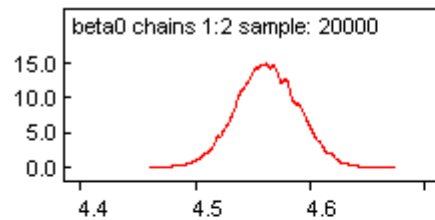
# priors:
beta0 ~ dnorm(0.0,1.0E-4)
beta1 ~ dnorm(0.0,1.0E-4)
tau.eps ~ dgamma(1.0E-3,1.0E-3)
var.eps <- 1.0/tau.eps
tau.gamma0 ~ dgamma(1.0E-3,1.0E-3)
var.gamma0 <- 1.0 /tau.gamma0
tau.gamma1 ~ dgamma(1.0E-3,1.0E-3)
var.gamma1 <- 1.0 /tau.gamma1
}
```

Modèle à intercept et pente aléatoires

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.563	0.02717	8.228E-4	4.51	4.563	4.616	1001	20000
beta1	0.1163	0.01306	1.249E-4	0.09083	0.1162	0.1422	1001	20000
var.eps	0.04378	0.002333	3.456E-5	0.03946	0.04371	0.04853	1001	20000
var.gamma0	0.09119	0.01197	1.291E-4	0.07039	0.09028	0.1171	1001	20000
var.gamma1	0.005341	0.002552	1.071E-4	0.001142	0.00506	0.01108	1001	20000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
CD4t	-277.817	-448.450	170.633	-107.185
total	-277.817	-448.450	170.633	-107.185



Modèle à intercept et pente aléatoires :

SAS – proc MIXED

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.563	0.02717	8.228E-4	4.51	4.563	4.616	1001	20000
beta1	0.1163	0.01306	1.249E-4	0.09083	0.1162	0.1422	1001	20000
var.eps	0.04378	0.002333	3.456E-5	0.03946	0.04371	0.04853	1001	20000
var.gamma0	0.09119	0.01197	1.291E-4	0.07039	0.09028	0.1171	1001	20000
var.gamma1	0.005341	0.002552	1.071E-4	0.001142	0.00506	0.01108	1001	20000

Covariance Parameter Estimates

Parm cov	Sujet	Estimation	Erreur type	Valeur Z	Pr > Z
Intercept	NUM_PAT	0.08896	0.01161	7.66	<.0001
T	NUM_PAT	0.005623	0.002622	2.14	0.0160
Residual		0.04338	0.002325	18.66	<.0001

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	4.5618	0.02725	148	167.42	<.0001
T	0.1164	0.01307	147	8.90	<.0001

Modèle à intercept et pente aléatoires avec matrice de variance-covariance structurée

$$\left\{ \begin{array}{l} y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij} \\ \gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\gamma_0}^2 & \sigma_{\gamma_0\gamma_1} \\ \sigma_{\gamma_0\gamma_1} & \sigma_{\gamma_1}^2 \end{bmatrix} \right) \\ \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2) \\ (\gamma_1 \dots \gamma_n) \perp (\varepsilon_1 \dots \varepsilon_n) \\ y_{ij} = \sqrt[4]{CD4_{ij}} \end{array} \right.$$

Modèle à intercept et pente aléatoires avec matrice de variance-covariance structurée

```
# Definition du modele
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      CD4t[i, j] ~ dnorm(mu[i,j], tau.eps);
      mu[i,j] <- beta0 + beta1*tps[i,j] +gamma[i,1]+gamma[i,2]*tps[i,j]
    }
    gamma[i,1:2] ~ dnorm(null[1:2], tau.gamma[,])
  }
}

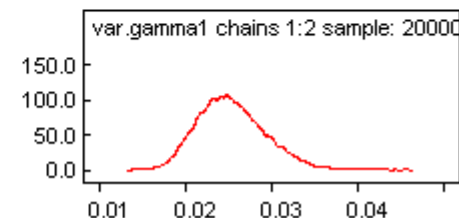
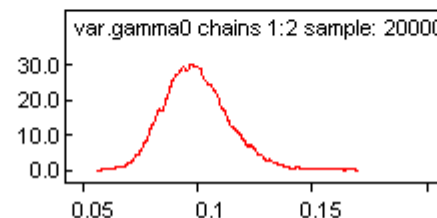
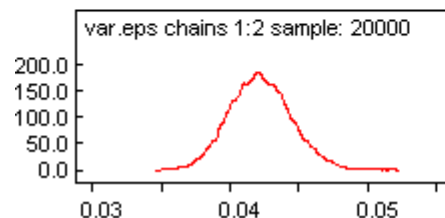
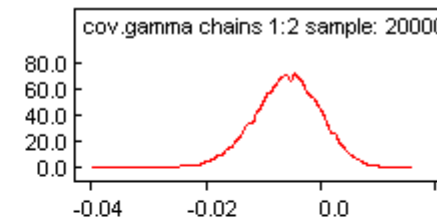
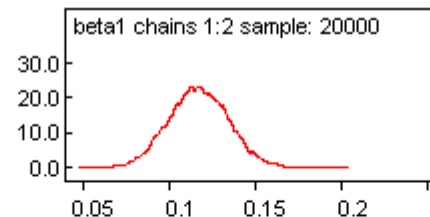
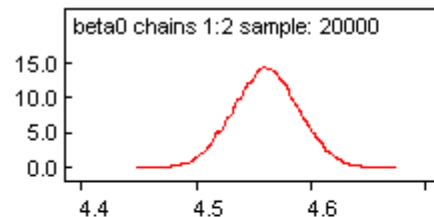
# priors:
beta0 ~ dnorm(0.0, 1.0E-4)
beta1 ~ dnorm(0.0, 1.0E-4)
tau.eps ~ dgamma(1.0E-3, 1.0E-3)
sigma.eps <- 1.0/tau.eps
tau.gamma[1:2, 1:2] ~ dwish(Omega[,], 2)
var.gamma[1:2, 1:2] <- inverse(tau.gamma[,])
var.gamma0 <- var.gamma[1, 1]
var.gamma1 <- var.gamma[2, 2]
cov.gamma <- var.gamma[1, 2]
}
```


Modèle à intercept et pente aléatoires avec matrice de variance-covariance structurée

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.561	0.02848	9.339E-4	4.505	4.561	4.617	1001	20000
beta1	0.1172	0.01736	3.066E-4	0.0835	0.1171	0.1515	1001	20000
cov.gamma	-0.005908	0.005861	7.034E-5	-0.01802	-0.005727	0.005013	1001	20000
var.eps	0.04222	0.002225	2.019E-5	0.03807	0.04216	0.04679	1001	20000
var.gamma0	0.09947	0.0139	1.467E-4	0.07528	0.09849	0.1294	1001	20000
var.gamma1	0.02528	0.00395	4.853E-5	0.01856	0.02493	0.0338	1001	20000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
CD4t	-312.047	-532.466	220.419	-91.628
total	-312.047	-532.466	220.419	-91.628



Modèle à intercept et pente aléatoires :

SAS - proc MIXED

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.561	0.02848	9.339E-4	4.505	4.561	4.617	1001	20000
beta1	0.1172	0.01736	3.066E-4	0.0835	0.1171	0.1515	1001	20000
cov.gamma	-0.005908	0.005861	7.034E-5	-0.01802	-0.005727	0.005013	1001	20000
var.eps	0.04222	0.002225	2.019E-5	0.03807	0.04216	0.04679	1001	20000
var.gamma0	0.09947	0.0139	1.467E-4	0.07528	0.09849	0.1294	1001	20000
var.gamma1	0.02528	0.00395	4.853E-5	0.01856	0.02493	0.0338	1001	20000

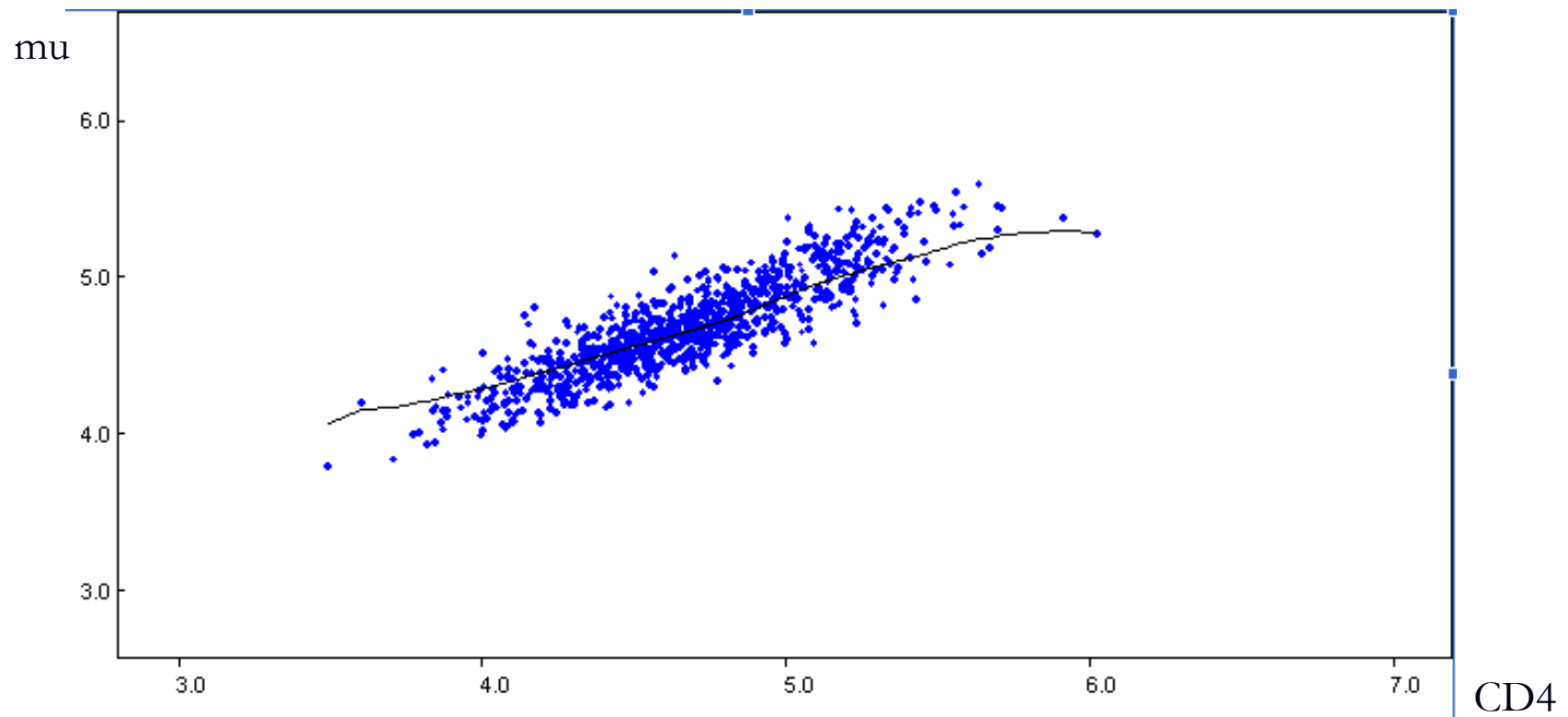
Covariance Parameter Estimates

Parm cov	Sujet	Estimation	Erreur type	Valeur Z	Pr > Z
UN(1,1)	NUM_PAT	0.08631	0.01261	6.84	<.0001
UN(2,1)	NUM_PAT	0.002197	0.004566	0.48	0.6304
UN(2,2)	NUM_PAT	0.005011	0.002869	1.75	0.0403
Residual		0.04358	0.002382	18.30	<.0001

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	4.5617	0.02693	148	169.41	<.0001
T	0.1164	0.01293	147	9.00	<.0001

Modèle à intercept et pente aléatoires : Courbes



Prise en compte d'un effet traitement (AZT+3TC=référence)

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + (\beta_2 + \beta_3 t_{ij}) 1_{trt=d4t+ddl} + (\beta_4 + \beta_5 t_{ij}) 1_{trt=switch} + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij}$$

$$\gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\gamma_0}^2 & \sigma_{\gamma_0\gamma_1} \\ \sigma_{\gamma_0\gamma_1} & \sigma_{\gamma_1}^2 \end{bmatrix} \right)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$(\gamma_1 \dots \gamma_n) \perp (\varepsilon_1 \dots \varepsilon_n)$$

$$y_{ij} = \sqrt[4]{CD4_{ij}}$$

Prise en compte d'un effet traitement (AZT+3TC=référence)

```
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      ...
      mu[i,j] <- beta0+beta1*tps[i,j]
                + (beta2+beta3*tps[i,j])*equals(grp[i],1) ← d4T+ddI
                + (beta4+beta5*tps[i,j])*equals(grp[i],2) ← Switch
                + gamma[i,1] + gamma[i,2]*tps[i,j]
    }
    ...
  }

# priors:
  ...
  beta2 ~ dnorm(0.0,1.0E-4)
  beta3 ~ dnorm(0.0,1.0E-4)
  beta4 ~ dnorm(0.0,1.0E-4)
  beta5 ~ dnorm(0.0,1.0E-4)
  ...
}
```

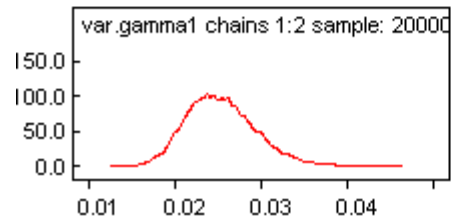
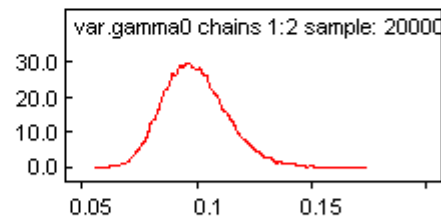
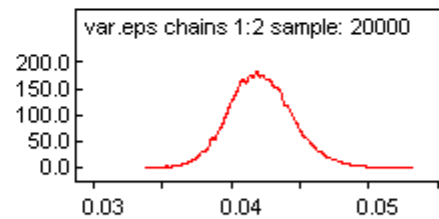
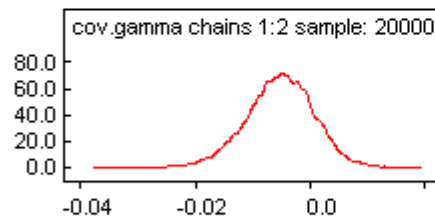
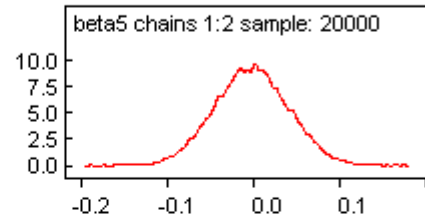
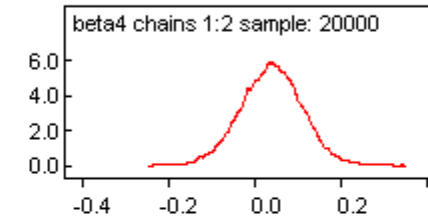
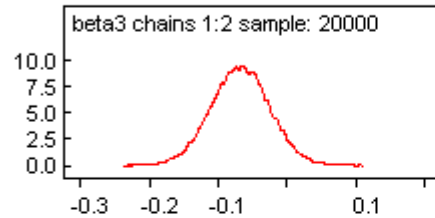
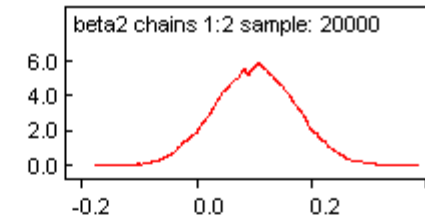
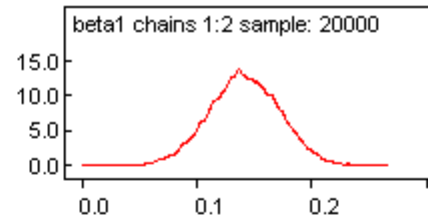
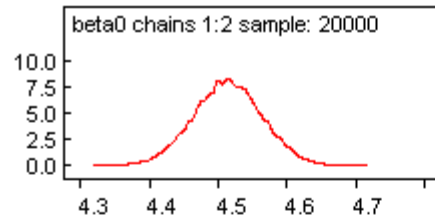
Prise en compte d'un effet traitement (AZT+3TC=référence)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	4.515	0.049	7.086E-4	4.418	4.515	4.611	1001	100000
beta1	0.1405	0.03041	2.404E-4	0.08146	0.1404	0.2003	1001	100000
beta2	0.1004	0.06941	0.001044	-0.0353	0.1003	0.2368	1001	100000
beta3	-0.06654	0.04271	3.298E-4	-0.1507	-0.06644	0.01701	1001	100000
beta4	0.03725	0.06947	0.001012	-0.09997	0.03754	0.1731	1001	100000
beta5	-0.001927	0.04282	3.127E-4	-0.08597	-0.00189	0.08178	1001	100000
cov.gamma	-0.005161	0.005853	3.342E-5	-0.01721	-0.004923	0.005785	1001	100000
var.eps	0.04217	0.002233	8.946E-6	0.038	0.04209	0.04679	1001	100000
var.gamma0	0.09943	0.01395	6.808E-5	0.07518	0.09835	0.1299	1001	100000
var.gamma1	0.02526	0.004	2.119E-5	0.0184	0.02493	0.03396	1001	100000

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
CD4t	-313.107	-534.339	221.231	-91.876
total	-313.107	-534.339	221.231	-91.876

Prise en compte d'un effet traitement (AZT+3TC=référence)



Prise en compte d'un effet traitement :

SAS - proc MIXED

node	mean	sd	2.5%	median	97.5%
beta0	4.515	0.049	4.418	4.515	4.611
beta1	0.1405	0.03041	0.08146	0.1404	0.2003
beta2	0.1004	0.06941	-0.0353	0.1003	0.2368
beta3	-0.06654	0.04271	-0.1507	-0.06644	0.01701
beta4	0.03725	0.06947	-0.09997	0.03754	0.1731
beta5	-0.001927	0.04282	-0.08597	-0.00189	0.08178

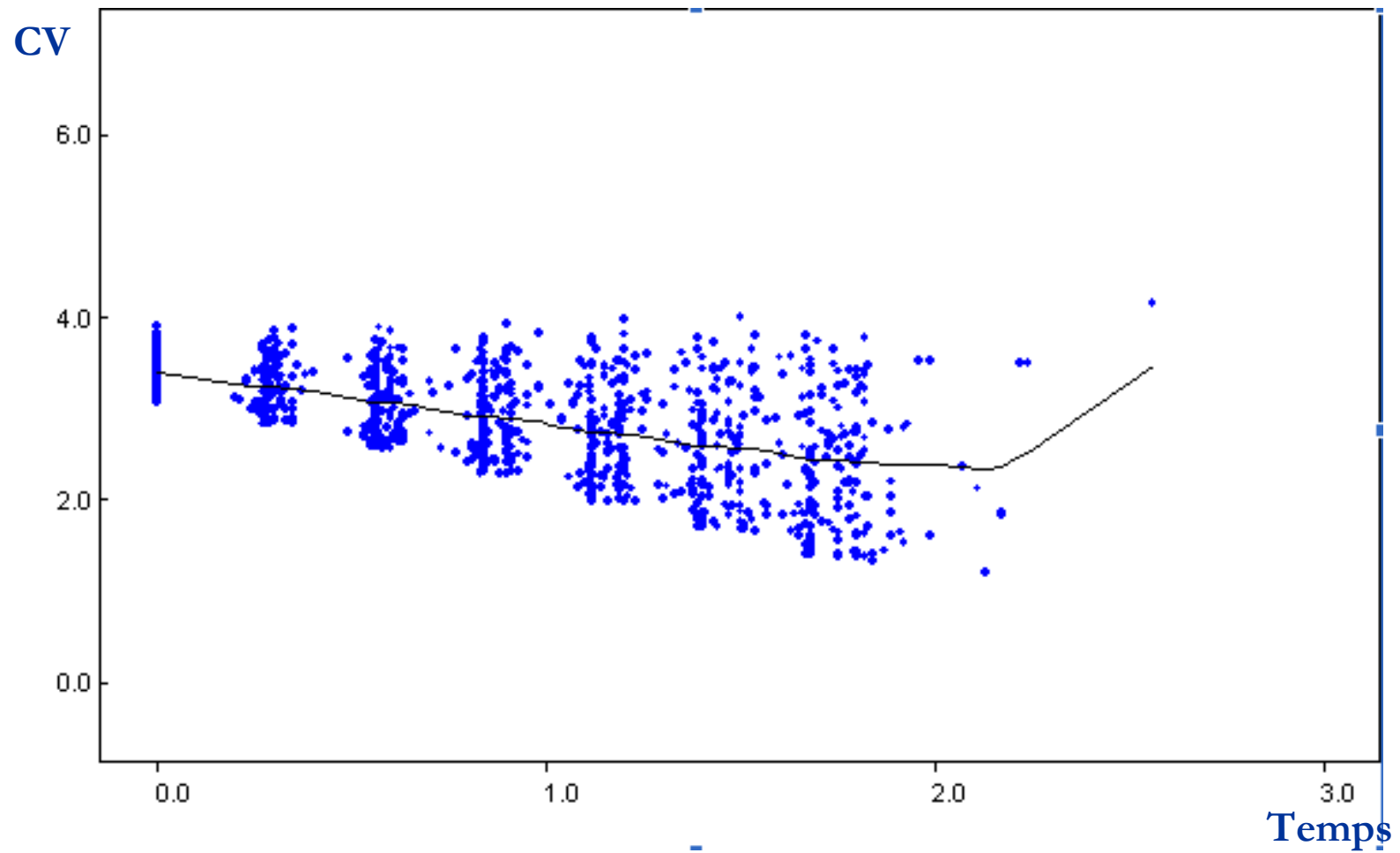
Solution for Fixed Effects

Effect	RAN_GRP	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		4.5158	0.04615	146	97.84	<.0001
T		0.1385	0.02200	145	6.30	<.0001
RAN_GRP	1	0.1021	0.06517	653	1.57	0.1177
RAN_GRP	2	0.03446	0.06552	653	0.53	0.5991
RAN_GRP	3	0
T*RAN_GRP	1	-0.06641	0.03091	653	-2.15	0.0321
T*RAN_GRP	2	0.001062	0.03107	653	0.03	0.9728
T*RAN_GRP	3	0

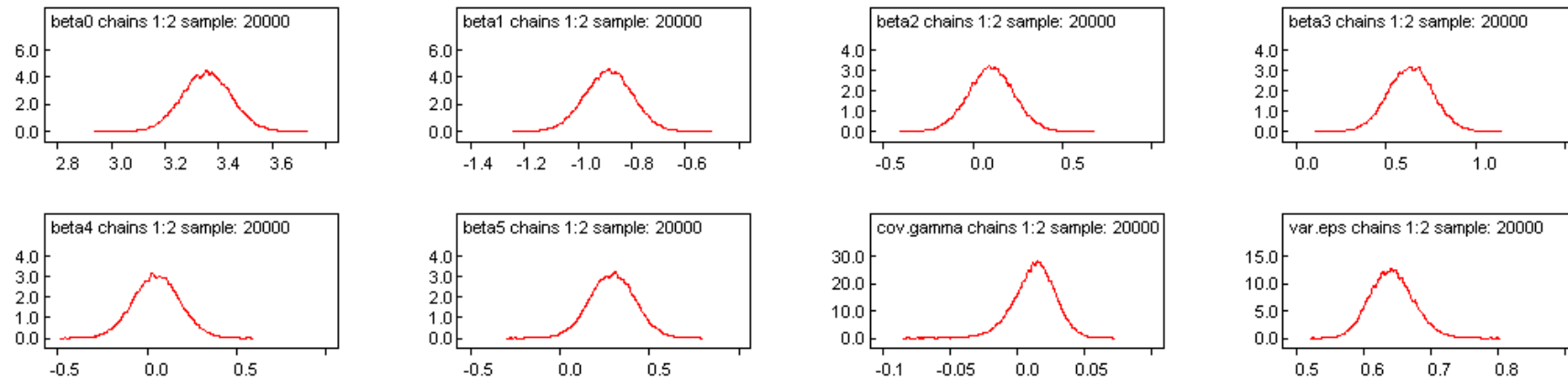
Modèle similaire pour la **CHARGE VIRALE**

$$\left\{ \begin{array}{l} y_{ij} = \beta_0 + \beta_1 t_{ij} + (\beta_2 + \beta_3 t_{ij}) 1_{trt=d4t+ddl} + (\beta_4 + \beta_5 t_{ij}) 1_{trt=switch} + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij} \\ \gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\gamma_0}^2 & \sigma_{\gamma_0\gamma_1} \\ \sigma_{\gamma_0\gamma_1} & \sigma_{\gamma_1}^2 \end{bmatrix} \right) \\ \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \\ (\gamma_1 \dots \gamma_n) \perp (\varepsilon_1 \dots \varepsilon_n) \\ y_{ij} = \log_{10}(CV_{ij}) \end{array} \right.$$

Modèle similaire pour la CHARGE VIRALE



Modèle similaire pour la CHARGE VIRALE



node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	3.356	0.09188	0.001033	3.178	3.355	3.536	1001	20000
beta1	-0.8855	0.08941	0.001003	-1.062	-0.885	-0.7126	1001	20000
beta2	0.1014	0.1278	0.001351	-0.1506	0.1011	0.3522	1001	20000
beta3	0.6373	0.1255	0.001393	0.3912	0.6381	0.882	1001	20000
beta4	0.05279	0.1295	0.001239	-0.2019	0.05283	0.3068	1001	20000
beta5	0.2928	0.1257	0.001403	0.04548	0.2931	0.54	1001	20000
cov.gamma	0.01235	0.01563	3.197E-4	-0.02074	0.01327	0.04098	1001	20000
var.eps	0.6424	0.03219	2.647E-4	0.5825	0.6414	0.7087	1001	20000
var.gamma0	0.09362	0.0241	5.148E-4	0.05464	0.09065	0.1482	1001	20000
var.gamma1	0.1018	0.02548	5.732E-4	0.06042	0.09888	0.1593	1001	20000

Modèle similaire pour la CHARGE VIRALE avec censure

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + (\beta_2 + \beta_3 t_{ij}) 1_{trt=d4t+ddl} + (\beta_4 + \beta_5 t_{ij}) 1_{trt=switch} + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij}$$

$$\gamma_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\gamma_0}^2 & \sigma_{\gamma_0\gamma_1} \\ \sigma_{\gamma_0\gamma_1} & \sigma_{\gamma_1}^2 \end{bmatrix} \right)$$

$$\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$$

$$(\gamma_1 \dots \gamma_n) \perp (\varepsilon_1 \dots \varepsilon_n)$$

$$y_{ij} = \log_{10}(CV_{ij}) 1_{[1.69; +\infty]}$$

Modèle similaire pour la CHARGE VIRALE avec censure

```
model {
  for(i in 1 : N) {
    for(j in 1 : T) {
      CV[i, j] ~ dnorm(mu[i, j], tau.eps)I(1.699,)
      mu[i, j] <- beta1*tps[i, j]
                    + (beta2+beta3*tps[i, j])*(equals(grp[i], 1)
                    + (beta4+beta5*tps[i, j])*(equals(grp[i], 2)
                    + gamma[i, 1] + gamma[i, 2]*tps[i, j]

    }
    ...
  }

# priors:
...
eft1 ~ dnorm(0.0, 1.0E-4)
eft2 ~ dnorm(0.0, 1.0E-4)
...
}
```

Modèle similaire pour la CHARGE VIRALE

avec censure

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	3.35	0.09032	8.654E-4	3.173	3.35	3.527	1001	20000
beta1	-0.8361	0.08688	9.651E-4	-1.007	-0.8365	-0.6646	1001	20000
beta2	0.1043	0.1279	0.001337	-0.1476	0.1045	0.3509	1001	20000
beta3	0.6026	0.1236	0.001434	0.3586	0.6032	0.8444	1001	20000
beta4	0.05602	0.1291	0.001253	-0.1982	0.05651	0.3075	1001	20000
beta5	0.2589	0.1238	0.001404	0.01785	0.2584	0.5057	1001	20000
cov.gamma	0.01078	0.01504	2.974E-4	-0.0212	0.01169	0.03817	1001	20000
var.eps	0.6474	0.03266	2.78E-4	0.5861	0.6462	0.7144	1001	20000
var.gamma0	0.09356	0.02429	5.381E-4	0.05444	0.09067	0.1489	1001	20000
var.gamma1	0.09664	0.0238	4.574E-4	0.0576	0.09423	0.1509	1001	20000

Partie IV
Diagnostiques de
convergence
CODA

Diagnostic de convergence (CODA)

http://www.people.fas.harvard.edu/~plam/teaching/methods/convergence/convergence_print.pdf

- **On reprend l'exemple** : Albi - Modèle à intercept et pente aléatoires avec matrice de variance-covariance structurée
- **Récupérer les fichiers Coda sous WinBUGS**
index.txt , chaine1.txt, chaine2.txt ...
- **Code R pour Coda**
`setwd(« monRepertoire \\ »)`
`library(coda)`
`codamenu()`

Diagnostic de convergence (CODA)

1: Output Analysis

2: Diagnostics

3: List/Change Options

4: Quit

1: Plots

2: Statistics

3: List/Change Options

4: Return to Main Menu

Diagnostic de convergence (CODA)

1: Output Analysis

2: **Diagnostics**

3: List/Change Options

4: Quit

1: Geweke

2: **Gelman and Rubin**

3: **Raftery and Lewis**

4: **Heidelberger and Welch**

5: **Autocorrelations**

6: **Cross-Correlations**

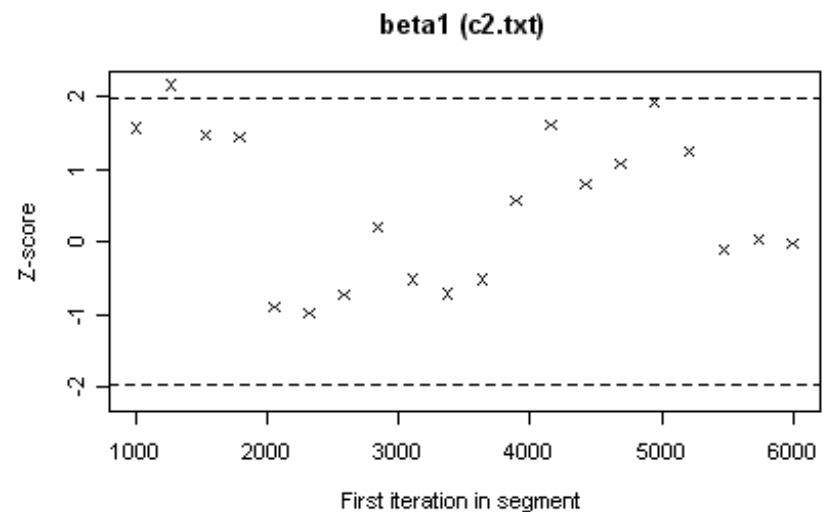
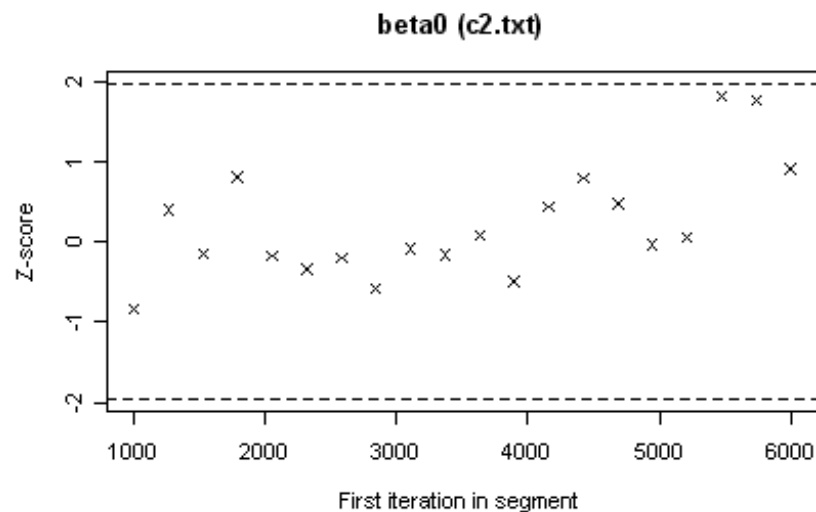
7: **List/Change Options**

8: **Return to Main Menu**

Diagnostic de convergence (CODA)

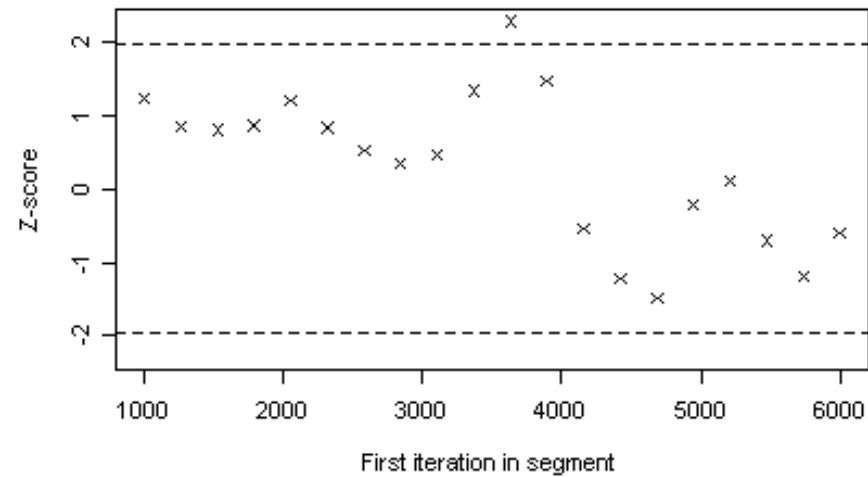
- **Test de Geweke** : teste l'égalité des moyennes entre la première partie de la chaîne (10%) et la fin de la chaîne (50%) grâce au Z-score.
- **Geweke-Brooks plot** : Graphique montrant les valeurs de la statistique de test du Z-score pour la comparaison des moyennes entre diverses répartitions.

$$Z = \frac{\mu_{fin} - \mu_{deb}^{(H_0)}}{\sigma} \sim N(0,1)$$

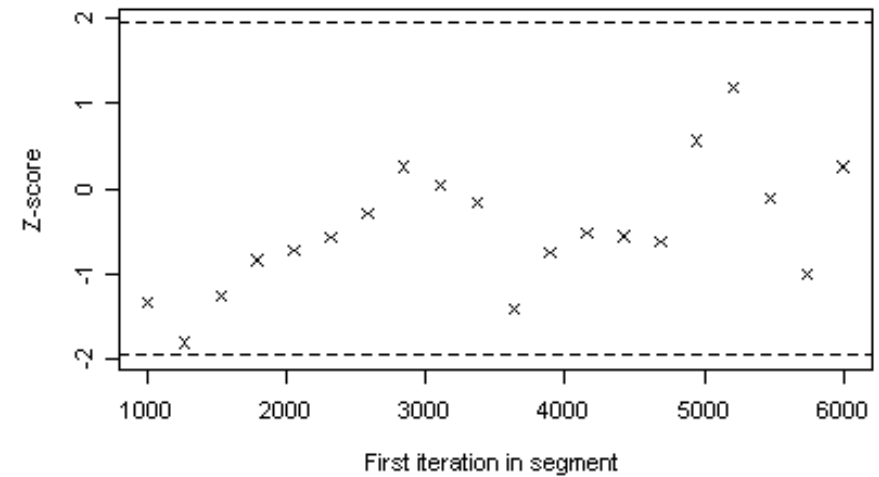


Diagnostic de convergence (CODA)

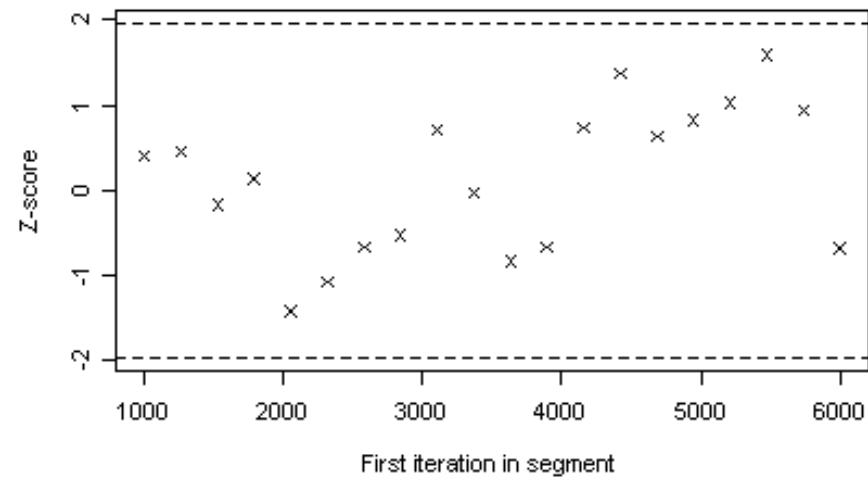
cov.gamma (c2.txt)



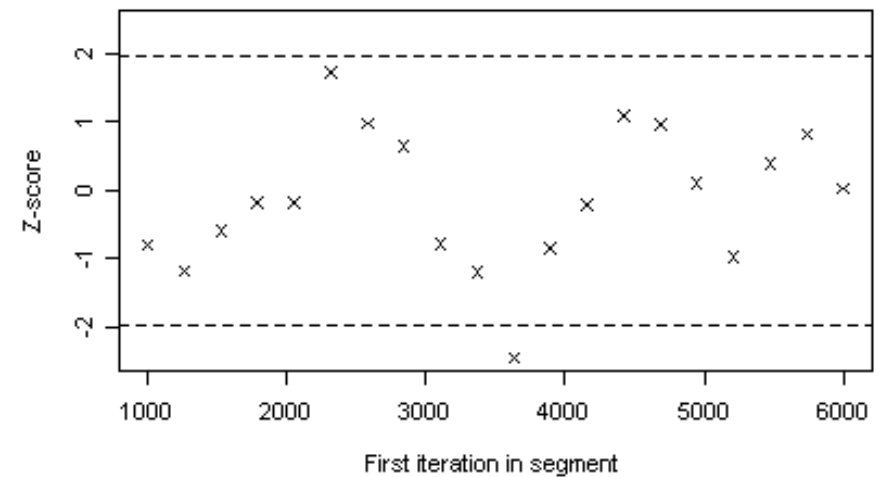
var.eps (c2.txt)



var.gamma0 (c2.txt)



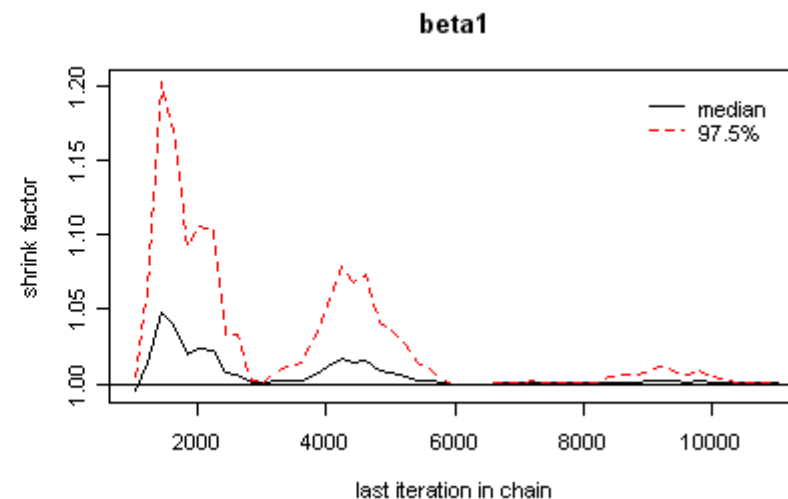
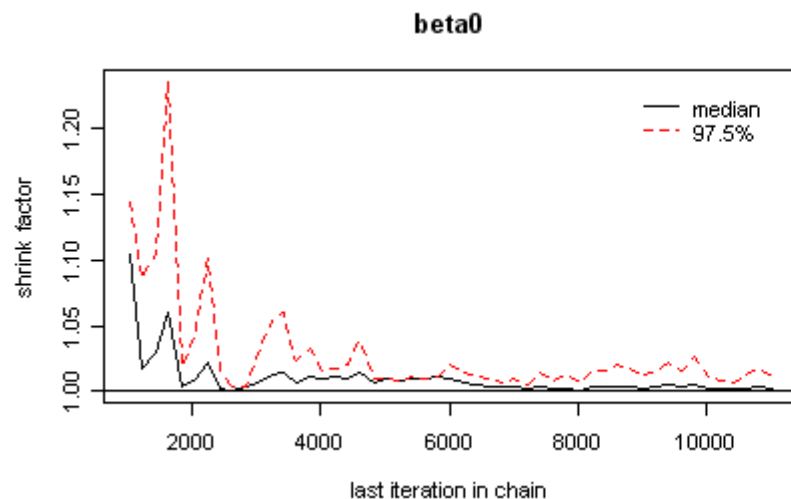
var.gamma1 (c2.txt)



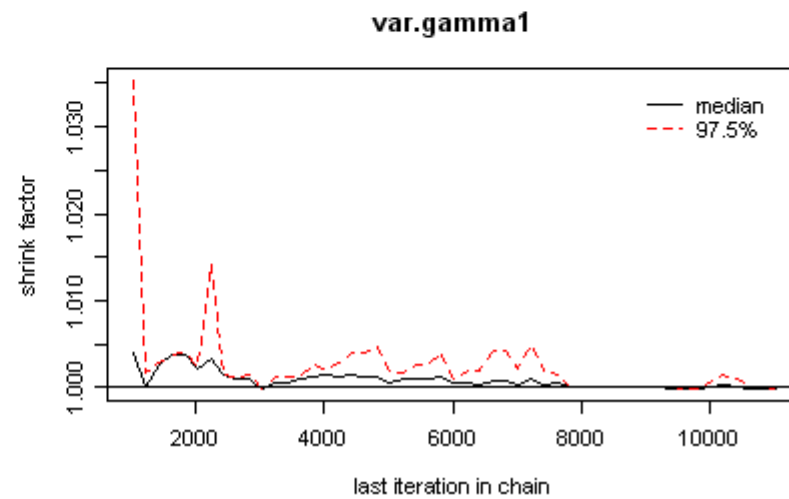
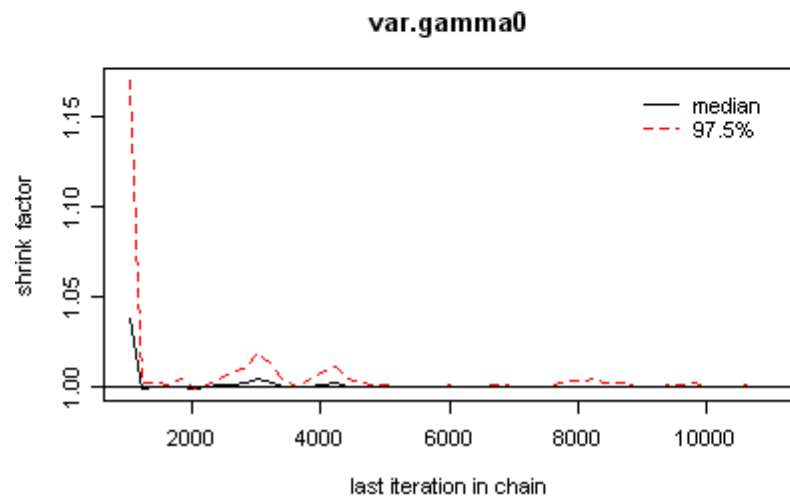
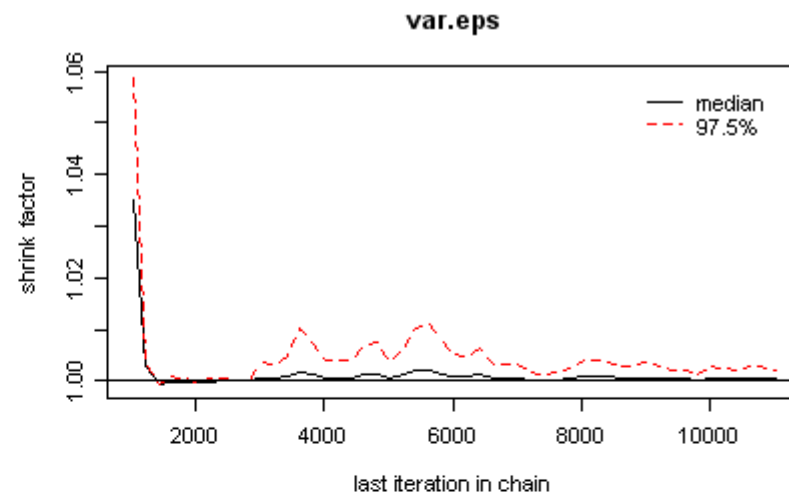
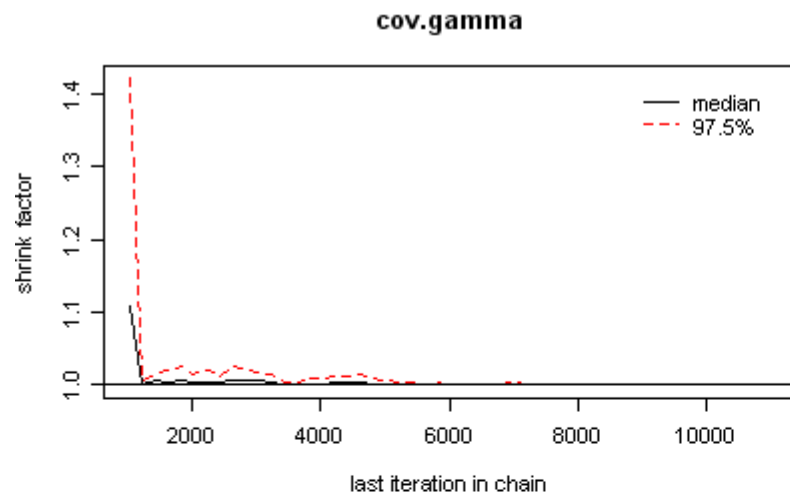
Diagnostic de convergence (CODA)

- **Gelman and Rubin** : Calcul du facteur d'échelle de réduction pour la chaîne moins les $n/2$ premiers éléments. Evalue la stationnarité (doit être inférieur à 1.1-1.2)

$$\text{shrink factor} = \frac{\text{Estimated variance of parameter}}{\text{Within variance of the chain for the parameter}}$$



Diagnostic de convergence (CODA)



Diagnostic de convergence (CODA)

- **Raftery and Lewis:** Teste le nombre d'itérations de chauffe et d'échantillonnage nécessaire pour avoir une estimation d'un quantile q avec une tolérance r .

M : Nombre d'itération en phase de chauffe

N : Taille du "pilot sampler" à générer

Nmin : Nombre d'itération en phase d'échantillonnage

I : Facteur d'indépendance, augmentation proportionnelle

`%c2.txt`

des itération attribuable à la dépendance dans la série.

```
Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
beta0	21	23901	3746	6.38
beta1	6	6462	3746	1.73
cov.gamma	6	7416	3746	1.98
var.eps	3	4129	3746	1.10
var.gamma0	3	4446	3746	1.19
var.gamma1	4	5192	3746	1.39

I>5 : « inquiétant ».
Montre:

- mauvais point de départ
- pauvre mélangeance
- grande corrélations

Diagnostic de convergence (CODA)

- **Heidelberger and Welch** : Test statistique de stationnarité de la chaîne de Markov.

Etape 1: Test itératif pour savoir si la chaîne est issue d'une loi stationnaire.

Etape 2: Halfwidth test, vérification de la taille du ratio entre la moitié de la longueur de l'intervalle de crédibilité à $(1-\alpha)\%$ et la moyenne a posteriori.

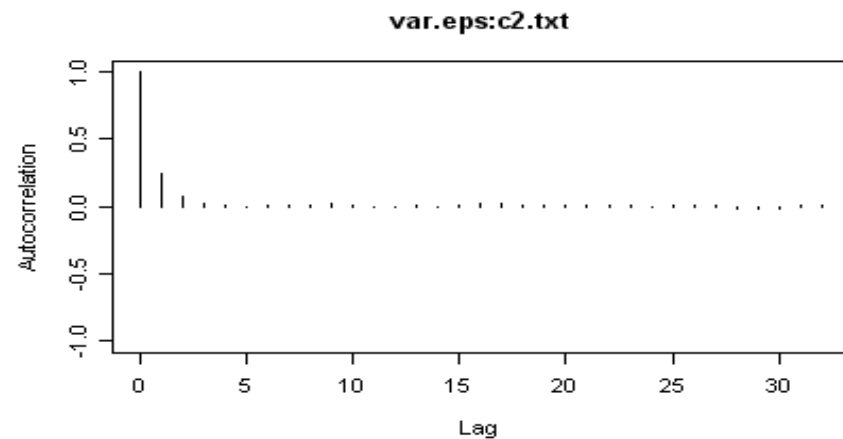
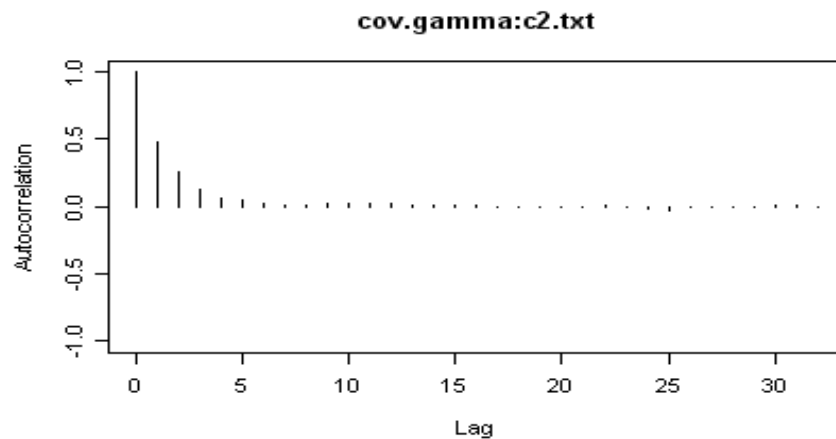
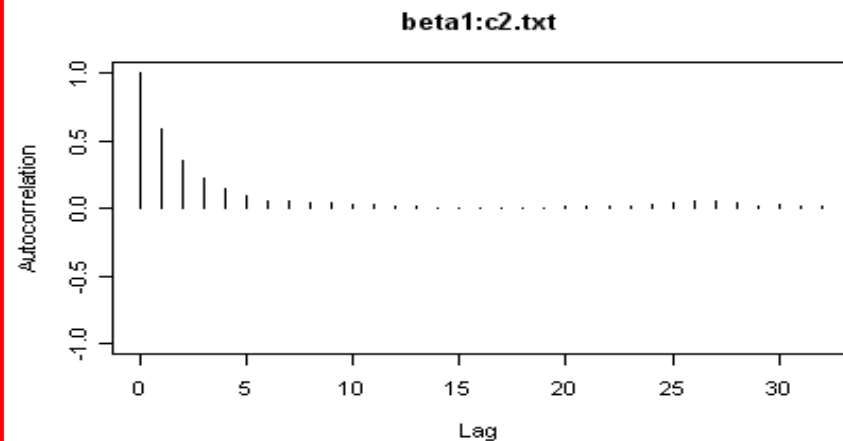
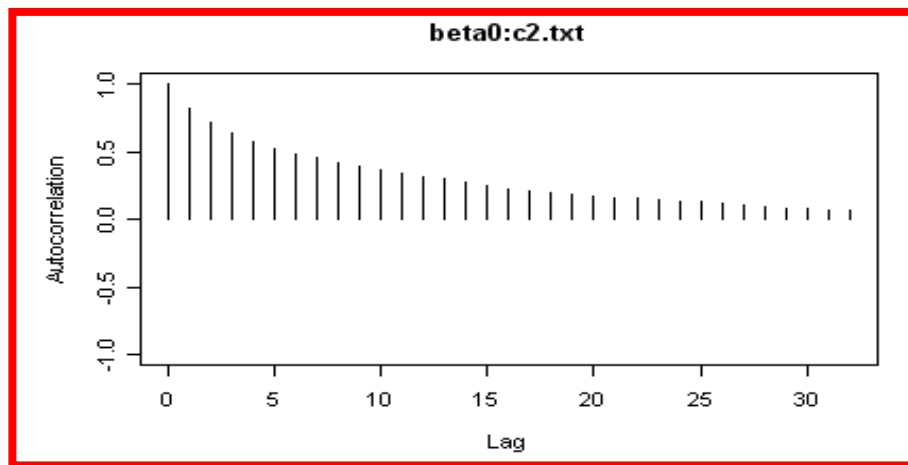
```
$c2.txt
```

```
Stationarity start      p-value
test          iteration
beta0         passed     1          0.984
beta1         passed     1          0.707
cov.gamma    passed     1          0.146
var.eps       passed     1          0.350
var.gamma0    passed     1          0.724
var.gamma1    passed     1          0.766
```

```
Halfwidth Mean      Halfwidth
test
beta0         passed     4.56181    2.53e-03
beta1         passed     0.11728    8.93e-04
cov.gamma    passed     -0.00583   1.74e-04
var.eps       passed     0.04227    5.71e-05
var.gamma0    passed     0.09944    3.54e-04
var.gamma1    passed     0.02526    1.32e-04
```


Diagnostic de convergence (CODA)

- **Autocorrélation** : Une grande autocorrélation indique un mélange lent et donc une convergence de la chaîne vers la distribution invariante lente.



Diagnostic de convergence (CODA)

- Changer l'épaisseur d'échantillonnage peut être une solution pour réduire l'autocorrélation sans forcément réduire l'information..

1: Output Analysis

2: Diagnostics

3: List/Change Options

4: Quit

1: List current options

2: Data Options

3: Plot Options

4: Summary Statistics Options

5: Diagnostics Options

6: Output Analysis

7: Diagnostics

8: Main Menu

1: List current data options

2: Select variables for analysis

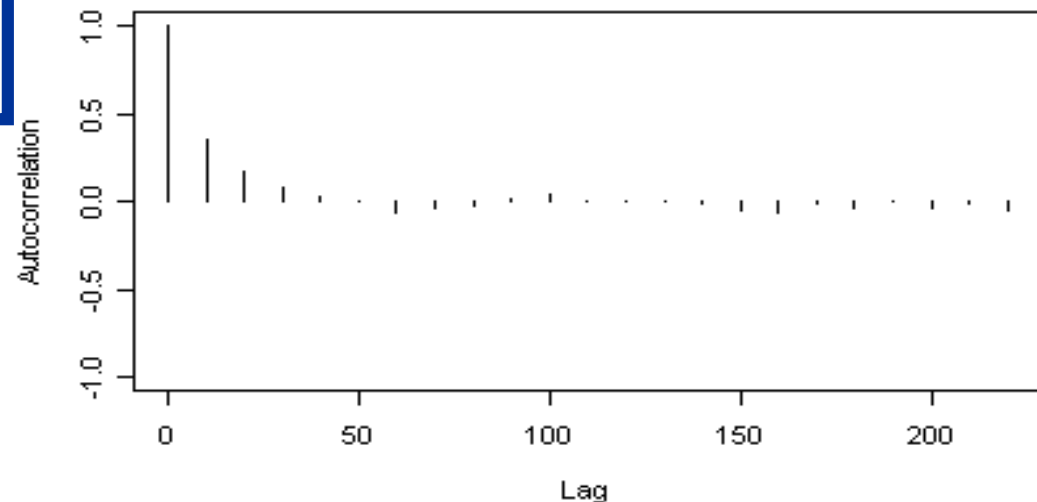
3: Select chains for analysis

4: Select iterations for analysis

5: Select thinning interval (METTRE A 10)

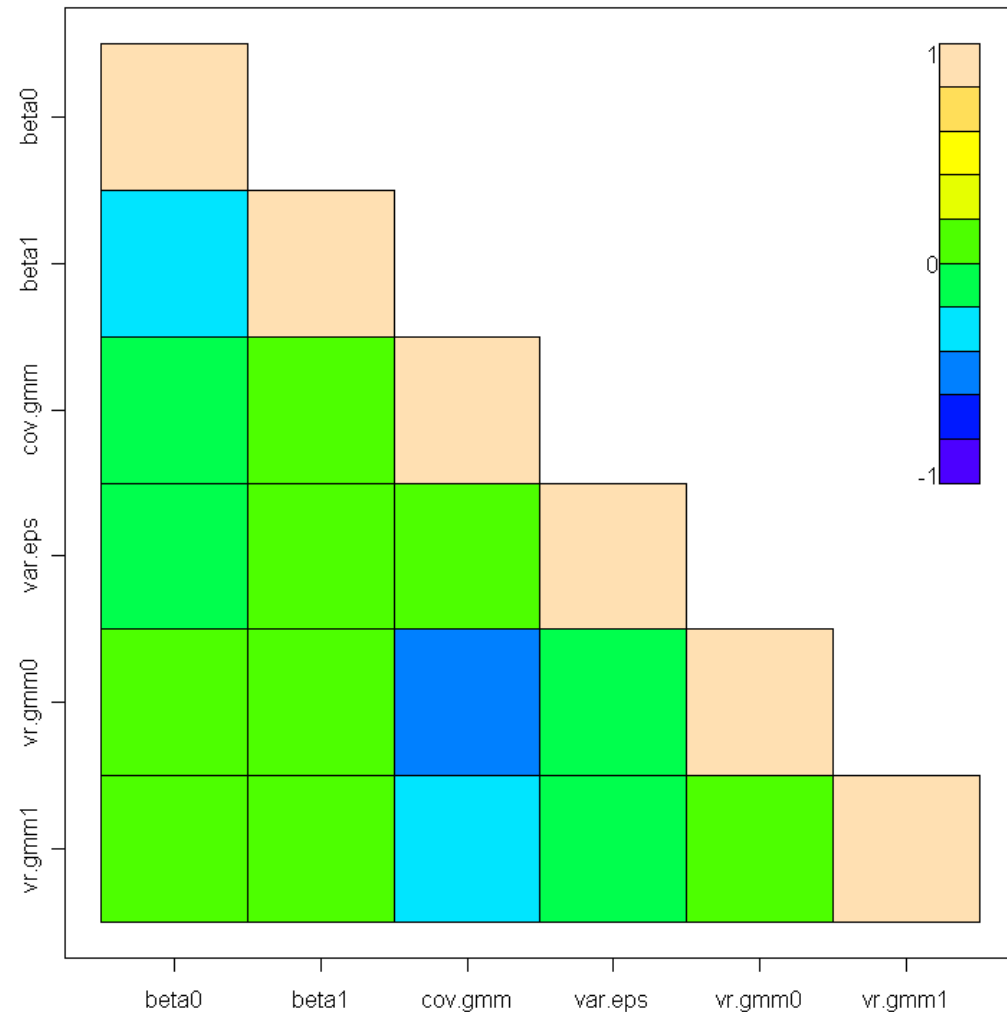
6: Return to main options menu

beta0:c2.txt



Diagnostic de convergence (CODA)

- **Cross Corrélation** :Renvoie la matrice de corrélation des variables estimées par les chaîne MCMC.



Partie V
Construction de
Doodle (DAG)

Doodles

- Les Doodles Winbugs sont des DAGs, **Directed Acyclic Graphs**. Ils permettent de **représenter graphiquement** un système d'équation de définition d'un modèle.
- Leur dessin peut paraître plus simple que la spécification par programmation du modèle. (**Write Code**)
- **Comment faire ? (Menu Doodle / Help->Doodle help)**
 - **Assiette** : ctrl + clic gauche
 - **Nœud** : clic gauche (type de nœud (stochastique, logique ou constant), type de loi, censure ...)
 - **Arrête** : Sélectionner le nœud cible. Appuyer sur contrôle, cliquez sur le nœud de provenance de la flèche. (idem pour supprimer)
 - **Suppression** : ctrl+backspace

Rats: a normal hierarchical model

$$Y_{ij} \sim \text{Normal}(\alpha_i + \beta_i(x_j - x_{\text{bar}}), \tau_c)$$

$$\alpha_i \sim \text{Normal}(\alpha_c, \tau_\alpha)$$

$$\beta_i \sim \text{Normal}(\beta_c, \tau_\beta)$$

