

Bayesian Statistics

Other Models

Brani Vidakovic

Professor

School of Industrial and Systems Engineering



Before We Begin...

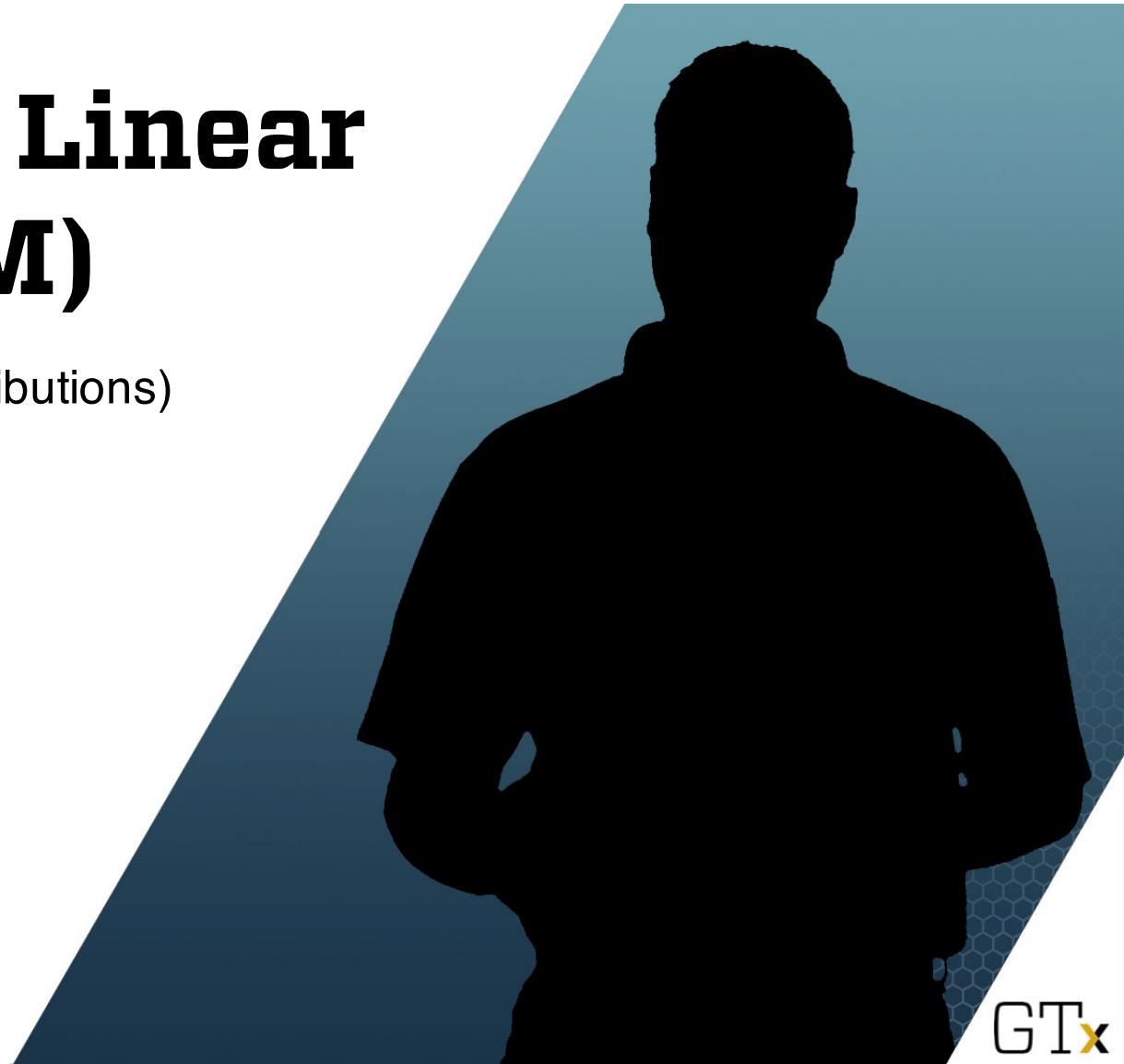
In this unit:

- GLMs
- Multinomial Regression
- Multilevel Models



Generalized Linear Models (GLM)

- Exponential (Family of Distributions)
- Logit, Probit, and CLogLog Regressions
- Poisson Regressions



GLM (Generalized Linear Models)

- What is generalized?
 - y_i remain independent, but distribution is generalized from normal to exponential family.
 - the mean of y_i is a function of $\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = l_i$
 $\mu_i = E y_i = g(l_i) \rightarrow g$ is a link function
 - Variance of y_i is not constant, depends on μ_i

Exponential Family

normal → exponential family

$$f(x|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}$$

θ – canonical parameter

- Normal is exponential family: $\theta = \mu$ since

$$f(x|\mu, \delta^2) = \exp \left\{ \frac{y\mu - \mu^2/2}{\delta^2} - \frac{1}{2} \left[\frac{y^2}{\delta^2} + \log(2\pi\delta^2) \right] \right\}$$

$\theta = \mu \rightarrow$ link is identity, $\mu = E y_i$

- [See the Exercises] $y_i \sim \text{Ber}(p)$;

$$\Rightarrow \theta = \log \frac{p}{1-p} \quad (= \text{logit } (p)), \quad p = E y_i$$

$$y_i \sim \text{Poi}(\lambda) \Rightarrow \theta = \log \lambda, \quad \lambda = E y_i$$

Typical link for binary (binomial) observations:

- logit: $\log \frac{p}{1-p} = F^{-1}(p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

 F is logistic cdf
- probit: $F^{-1}(p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, F is normal cdf.
- cloglog: complementary log-log, F^{-1} of cloglog
 F is cdf of Gumbel type I distribution
$$F(x) = 1 - \exp(-\exp(x))$$
$$F^{-1}(p) = \text{cloglog}(p) = \log(-\log(1-p)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

For Poisson regression link is log:

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

- Standard measures of model performance are deviances.
- $D = -2 \log \frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}}$
- Logistic: $D = -2 \sum_{i=1}^k y_i \log \frac{\hat{y}_i}{y_i} + (n_i - y_i) \log \frac{n_i - \hat{y}_i}{n_i - y_i}$
where $\hat{y}_i = n_i \hat{p}_i$ is the model fit for y_i
For saturated model: $\hat{p}_i = \frac{y_i}{n_i}$, $\hat{y}_i = y_i$
Here, n_i is the number of cases for which covariates coincide, and k is the number of such groups.
- For Poisson regression:

$$D = 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right)$$

where $\hat{y}_i = \exp\{b_0 + b_1 x_{i1} + \cdots + b_k x_{ik}\}$

 OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

 arrhythmia

 BUGS

ARRHYTHMIA

Patients who undergo Coronary Artery Bypass Graft Surgery (CABG) have an approximate 19-40% chance of developing atrial fibrillation (AF). AF can lead to blood clots forming causing greater in-hospital mortality, strokes, and longer hospital stays. While this can be prevented with drugs, it is very expensive and sometimes dangerous if not warranted. Ideally, several risk factors which would indicate an increased risk of developing AF in this population could save lives and money by indicating which patients need pharmacological intervention. Researchers began collecting data from CABG patients during their hospital stay such as demographics like age and sex, as well as heart rate, cholesterol, operation time, etc.. Then, the researchers recorded which patients developed AF during their hospital stay. Researchers now want to find those pieces of data which indicate high risk of AF. In the past, indicators like age, hypertension, and body surface area (BSA) have been good indicators, though these alone have not produced a satisfactory solution.

Fibrillation occurs when the heart muscle begins a quivering motion instead of a normal, healthy pumping rhythm. Fibrillation can affect the atrium (atrial fibrillation) or the ventricle (ventricular fibrillation); ventricular fibrillation is imminently life-threatening.

Atrial fibrillation is the quivering, chaotic motion in the upper chambers of the heart, known as the atria. Atrial fibrillation is often due to serious underlying medical conditions, and should be evaluated by a physician. It is not typically a medical emergency.

Ventricular fibrillation occurs in the ventricles (lower chambers) of the heart; it is always a medical emergency. If left untreated, ventricular fibrillation (VF, or V-fib) can lead to death within minutes. When a heart goes into V-fib, effective pumping of the blood stops. V-fib is

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

arrhythmia

the atria. Atrial fibrillation is often due to serious underlying medical conditions, and should be evaluated by a physician. It is not typically a medical emergency.

Ventricular fibrillation occurs in the ventricles (lower chambers) of the heart; it is always a medical emergency. If left untreated, ventricular fibrillation (VF, or V-fib) can lead to death within minutes. When a heart goes into V-fib, effective pumping of the blood stops. V-fib is considered a form of cardiac arrest, and an individual suffering from it will not survive unless cardiopulmonary resuscitation (CPR) and defibrillation are provided immediately.

DATA Arrhythmia (81 patient)

Y = Fibrillation

X1 = Age

X2 = Aortic Cross Clamp Time

X3 = Cardiopulmonary Bypass Time:
Bypass of the heart and lungs as, for example, in open heart surgery.
Blood returning to the heart is diverted through a heart-lung machine (a pump-oxygenator) before returning it to the arterial circulation.
The machine does the work both of the heart (pump blood) and the lungs (supply oxygen to red blood cells).

X4 = ICU Time
Intensive Care Unit

X5 = Avg Heart Rate

X6 = Left Ventricle Ejection Fraction

X7 = Hypertension|

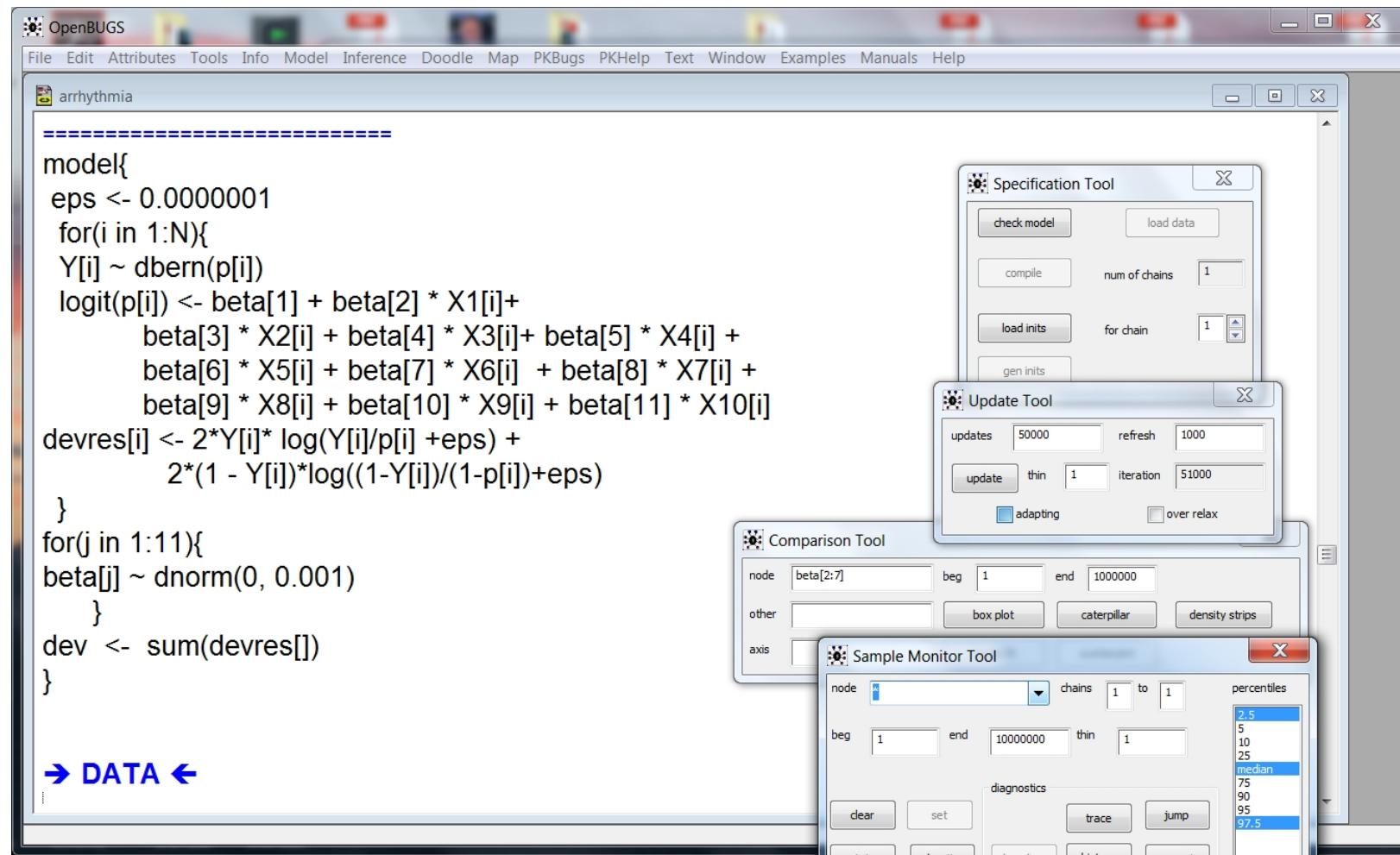
X8 = Gender [1 -Female; 0-Male]

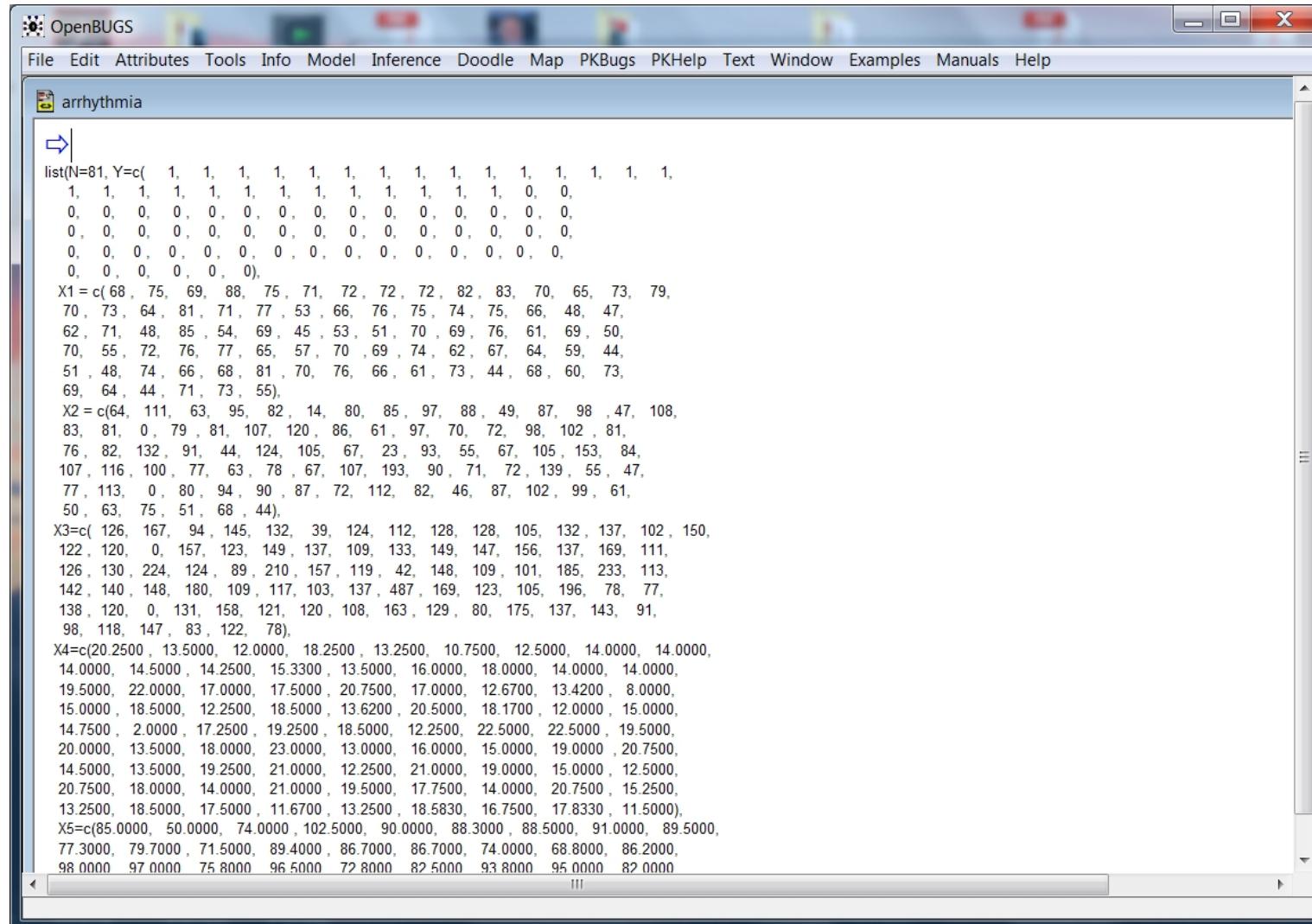
X9 = Diabetis

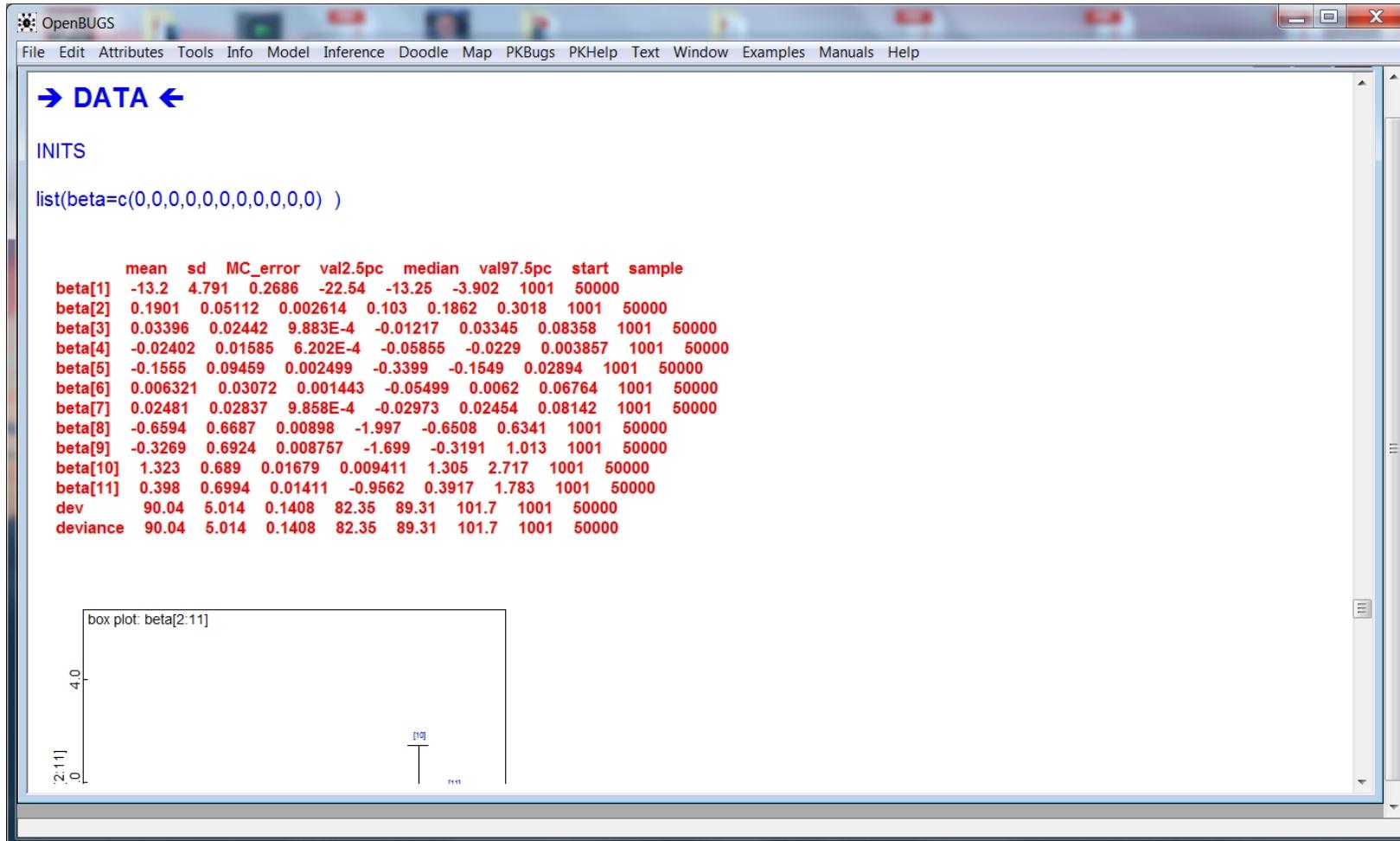
X10 = Previous MI

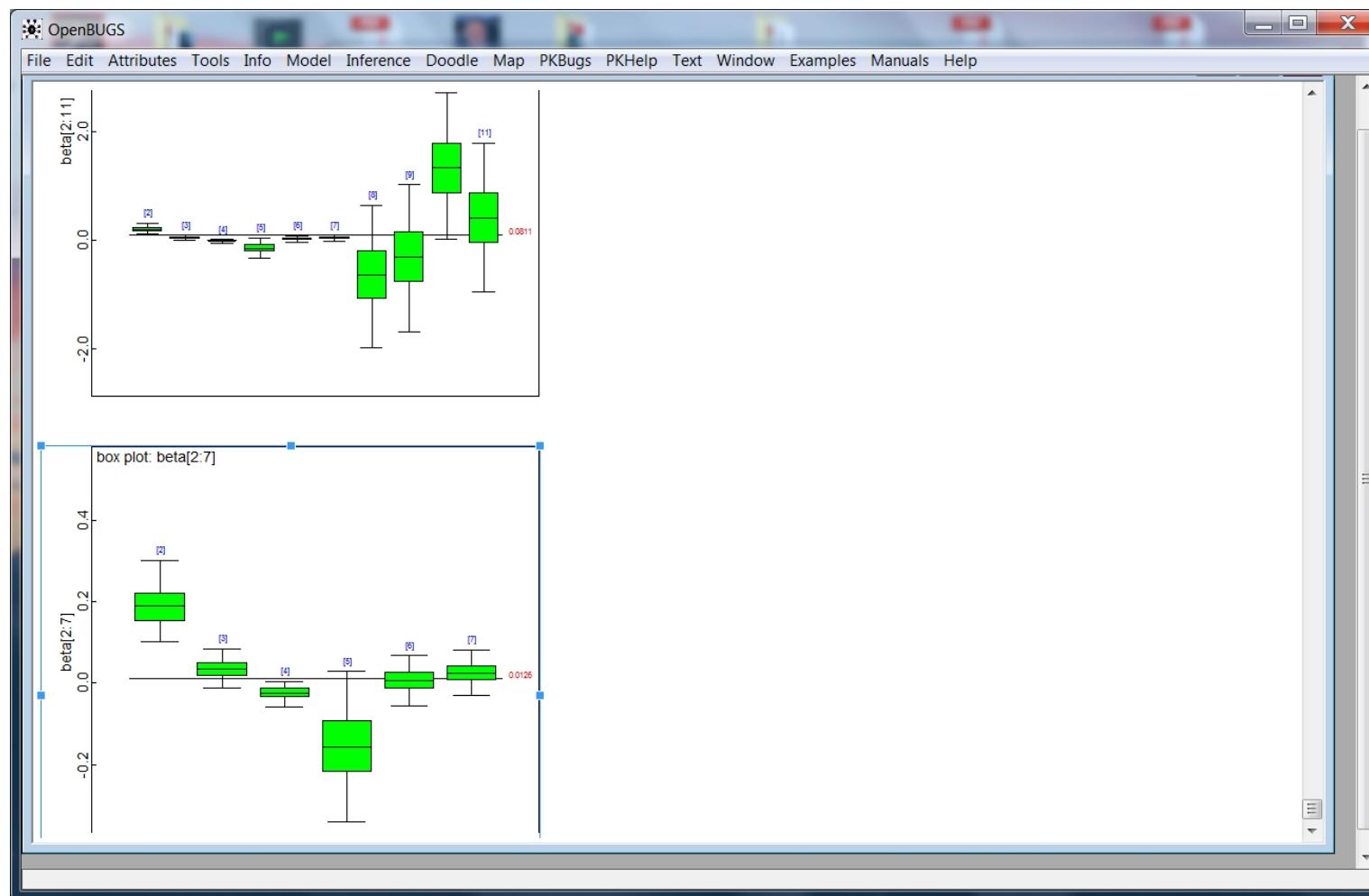
=====

OpenBUGS version 3.2.3 rev 1012









OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

ants

 ANTS

The data discussed in Gotelli and Ellison (2002), provide the ant species richness (number of ant species) found in 64-square-meter sampling grids in 22 forests (coded as 1) and 22 bogs (coded as 2) surrounding the forests in Connecticut, Massachusetts, and Vermont. The sites span 3° of latitude in New England. There are 44 observations on four variables (columns in data set): Ants – number of species, Habitat – forests (1) and bogs (2), Latitude, and Elevation – in meters above sea level.

(a) Using Poisson regression, model the number of ant species (Ants) with covariates Habitat and Elevation.

(b) For a sampling grid unit located in a forest at the elevation of 100 m how many species the model from (a) predicts? For the model coefficients and the prediction report 95% credible sets.

```
model{  
for( i in 1:n ) {  
ants[i] ~ dpois(lambda[i])  
lambda[i] <- exp( beta[1] + beta[2] * habitat[i] + beta[3] * elevation[i] )  
# or  
# log(lambda[i]) <- beta[1] + beta[2] * habitat[i] + beta[3] * elevation[i]  
}  
for (j in 1:3){  
beta[j] ~ dnorm(0, 0.0001)  
}  
lambdaStar <- exp( beta[1] + beta[2] + 100 * beta[3])  
#average number of species  
ystar ~ dpois(lambdaStar) #predicted number of species  
}
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

ants

DATA

list(n=44)

```
ants[] habitat[] elevation[]
6 1 389
16 1 8
18 1 152
17 1 1
9 1 210
15 1 78
7 1 47
12 1 491
14 1 121
9 1 95
10 1 274
10 1 335
4 1 543
5 1 323
7 1 158
7 1 313
4 1 468
6 1 362
6 1 236
8 1 30
6 1 353
6 1 133
5 2 389
6 2 8
14 2 152
7 2 1
4 2 210
8 2 78
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

ants

```
2 2 47
3 2 491
4 2 121
8 2 95
8 2 274
4 2 335
2 2 543
7 2 323
2 2 158
3 2 313
3 2 468
2 2 362
3 2 236
2 2 30
5 2 353
5 2 133
END
```

INITS

```
list(beta=c(0,0,0))
```

RESULTS

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	3.171	0.1814	0.006551	2.816	3.171	3.525	10001	100000
beta[2]	-0.6371	0.1156	0.003908	-0.8667	-0.6363	-0.4107	10001	100000
beta[3]	-0.001501	3.803E-4	1.238E-5	-0.002274	-0.001496	-7.885E-4	10001	100000
lambdastar	10.88	0.8463	0.0248	9.28	10.85	12.63	10001	100000
ystar	10.89	3.4	0.0265	5.0	11.0	18.0	10001	100000

Gotelli, N. J. and A. M. Ellison. 2002. Biogeography at a regional scale: determinants of ant species density in bogs and forests of New England. Ecology 83: 1604-1609.

Multinomial Logit

$$y_1, y_2, \dots, y_n \sim \text{Mn}(\mathbf{p}, 1)$$

$$\mathbf{p} = (p_1, p_2, \dots, p_K)$$

$$y_i = (y_{i1}, y_{i2}, \dots, y_{iK}), y_{ij} = 1, y_{i,j} = 0 \quad j \in \{1, \dots, K\}$$

For example: $y_i = (0, 0, 0, 1, 0)$

$$K = 5, y_{i4} = 1, y_{i,\neq 4} = 0, \mathbf{p} = (p_1, p_2, \dots, p_K)$$

○ ith subject: covariates $x_{i1}, \dots, x_{i,p-1}$

$$p_{ij} = \frac{\exp\{\eta_{ij}\}}{\sum_{k=1}^K \exp\{\eta_{ik}\}}, \quad \eta_{ij} = \beta_{0j} + \beta_{1j}x_{i1} + \dots + \beta_{p-1,j}x_{i,p-1}$$

Multinomial Logit (cont.)

- $p_{i1} = P(y_{i1} = 1) = \frac{1}{1 + \sum_{k=2}^K \exp\{\eta_{ik}\}}$
- $p_{ij} = P(y_{ij} = 1) = \frac{\exp\{\eta_{ij}\}}{1 + \sum_{k=2}^K \exp\{\eta_{ik}\}}$

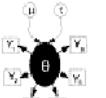
$$\log \frac{p_{ij}}{p_{i1}} = \eta_{ij} = \beta_{0j} + \beta_{1j}x_{i1} + \cdots + \beta_{p-1,j}x_{i,p-1}$$

- NHANESmulti.odc

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

NHANESmulti



BUGS Multinomial Regression

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations.

Assume that N subjects select a choice from K categories. The i-th subject is characterized by 3 covariates $x[i,1]$, $x[i,2]$, and $x[i,3]$. Given the covariates, model the probability of a subject selecting the category k, $k=1,\dots,K$.

```
model {
for (i in 1:N) {
# multinomial likelihood
y[i,1:K] ~ dmulti(P[i,1:K],1)
for (j in 1:K) {
eta[i,j] <- exp(b[1,j]+ b[2,j]*X[i,1]+ b[3,j]*X[i,2]+b[4,j]*X[i,3])
P[i,j] <- eta[i,j]/sum(eta[i,1:K])}}
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

NHANESmulti

```
model {
for (i in 1:N) {
# multinomial likelihood
y[i,1:K] ~ dmulti(P[i,1:K],1)
for (j in 1:K) {
eta[i,j] <- exp(b[1,j]+ b[2,j]*X[i,1]+ b[3,j]*X[i,2]+b[4,j]*X[i,3])
P[i,j] <- eta[i,j]/sum(eta[i,1:K])}

for(p in 1:4){
b[p,1] <- 0
for (j in 2:K) {b[p, j] ~ dnorm(0,0.1)} }

# New subject covariates
X1 <- 3
X2 <- 3
X3 <- 30
for(j in 1:K) {
etastar[j] <- exp(b[1,j]+ b[2,j]*X1+ b[3,j]*X2+b[4,j]*X3)
pstar[j] <- etastar[j]/sum(etastar[1:K])}

}
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

NHANESmulti

DATA

```
list(N=10, K=5 )  
  
y[,1] y[,2] y[,3] y[,4] y[,5]  
1 0 0 0 0  
0 1 0 0 0  
1 0 0 0 0  
0 0 1 0 0  
0 1 0 0 0  
0 0 1 0 0  
0 0 0 1 0  
0 0 0 0 1  
0 0 0 0 1  
0 0 0 1 0  
END  
  
X[,1] X[,2] X[,3]  
2 4 9  
1 5 10  
1 6 14  
2 4 21  
2 4 22  
2 6 30  
?? ??
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

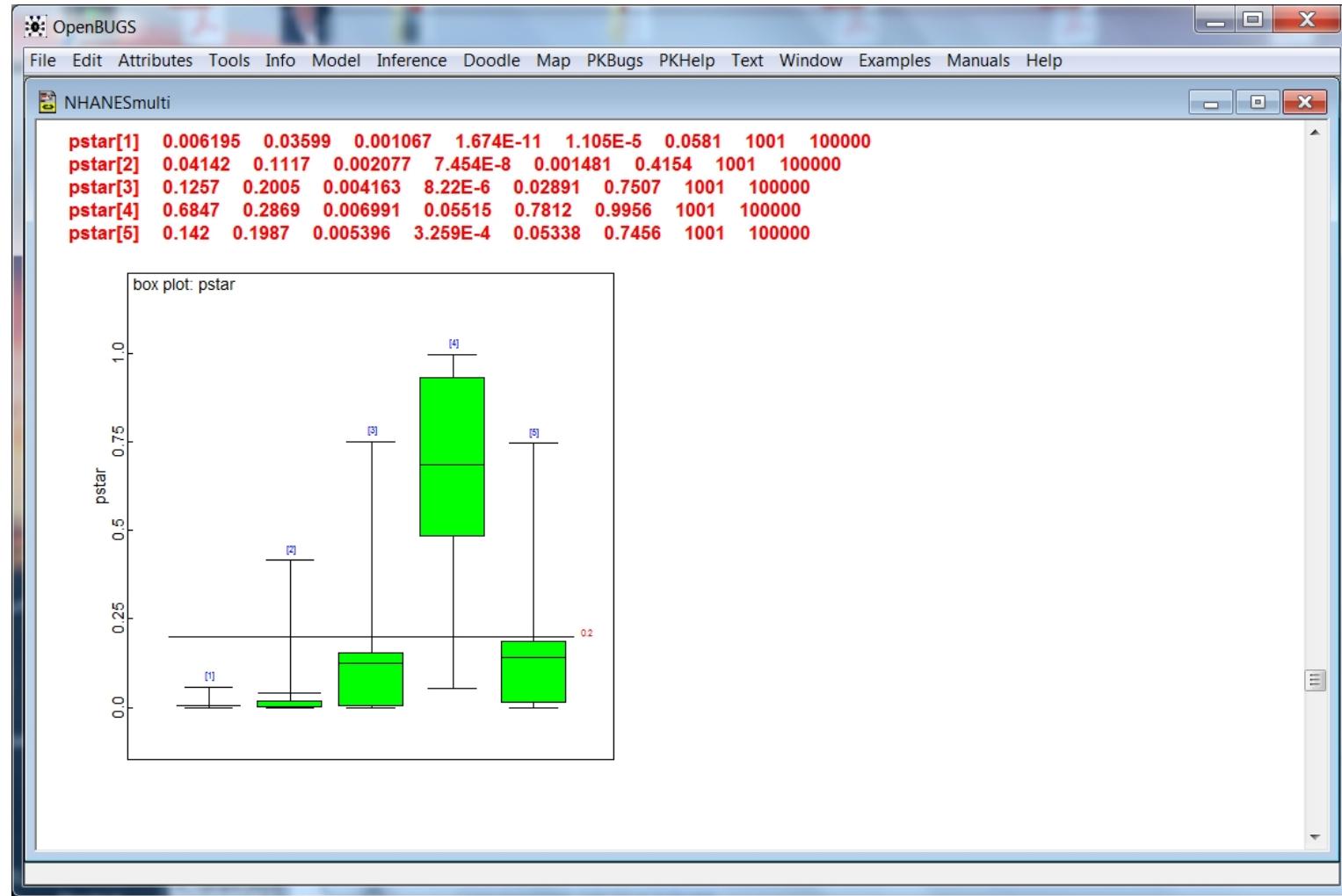
NHANESmulti

```
1 6 14
2 4 21
2 4 22
2 6 30
3 3 33
3 2 36
3 1 40
4 1 44
END

INITS

b[1] b[2] b[3] b[4] b[5]
NA 1 1 1 1
END

| mean sd MC_error val2.5pc median val97.5pc start sample
pstar[1] 0.006195 0.03599 0.001067 1.674E-11 1.105E-5 0.0581 1001 100000
pstar[2] 0.04142 0.1117 0.002077 7.454E-8 0.001481 0.4154 1001 100000
pstar[3] 0.1257 0.2005 0.004163 8.22E-6 0.02891 0.7507 1001 100000
pstar[4] 0.6847 0.2869 0.006991 0.05515 0.7812 0.9956 1001 100000
pstar[5] 0.142 0.1987 0.005396 3.259E-4 0.05338 0.7456 1001 100000
```



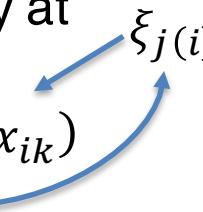
Multilevel Models

- Random effects models
- Mixed effects models

Example: $y_i = \alpha_{j(i)} + \beta x_i + \epsilon_i$, for students $i = 1, \dots, n$
 $\alpha_j = a + b\mu_j + \eta_j$ for schools $j = 1, \dots, J$

- Learning about treatment effects that vary.
- “Borrowing strength” – using all data to make inference about small groups.
- Getting better fitting models. Account for the uncertainty at different levels

Example: $y_i \sim \text{Poi}(\lambda_i)$ $\lambda_i = \log(\beta_{i0} + \beta_1 x_{i1} + \dots + \beta_k x_{ik})$
 $Ey_i = \text{Var } y_i = \lambda_i$ but $Ey_i < \text{Var } y_i$



paraguay

Paraguay Vaccination Status

This example considers factors influencing the vaccination status among 3424 children of 2552 mothers among 264 clusters in Paraguay. In this analysis, we're specifically interested in mother-level factors related to child immunization.

However, there is randomness associated with different clusters.

ID3: Cluster number
VACCODE: =1 if fully immunized, =0 otherwise
LB.TOT: No. of live births
MAGE2: mother age <20 =1, otherwise = 0
UN2: consensual union = 1, otherwise = 0
TOILET2: unsafe toilet in hh = 1, otherwise = 0
PR.SPOC1: spouse unskilled laborer = 1, otherwise = 0
SPANISH2: Spanish not hh language = 1, otherwise = 0

```
model
{
for( i in 1 : N ) {
vaccine[i] ~ dbern(p[i])
logit(p[i]) <- beta0 + beta1 *lb.tot[i] + beta2 *mage2[i] +beta3 * un2[i] +
beta4* toilet2[i] + beta5*pr.s poc1[i]+beta6*spanish2[i]+b[id3[i]]
}
b[1]<-0
for(j in 2:id3[N])
{
b[j] ~ dnorm(0.0, tau.b)
```

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

paraguay

```
b[1]<-0
for(j in 2:id3[N])
{
b[j] ~ dnorm(0.0, tau.b)
}
beta0 ~ dnorm(0.0,1.0E-6)
beta1 ~ dnorm(0.0,1.0E-6)
beta2 ~ dnorm(0.0,1.0E-6)
beta3 ~ dnorm(0.0,1.0E-6)
beta4 ~ dnorm(0.0,1.0E-6)
beta5 ~ dnorm(0.0,1.0E-6)
beta6 ~ dnorm(0.0,1.0E-6)
tau.b ~ dgamma(0.001,0.001)
sigma2.b <- 1/(tau.b)
sigma.b <- 1 / sqrt(tau.b)
}
```

DATA

```
⇒
list(vaccine = c(0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0,
1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1,
0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1,
0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1,
0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1,
1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1,
1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1,
0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
```



OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

paraguaynocluster

```
model{
for( i in 1 : N ) {
vaccode[i] ~ dbern(p[i])
logit(p[i]) <- beta[1] + beta[2] *lb.tot[i] + beta[3] *mage2[i] +beta[4] * un2[i] +
beta[5]* toilet2[i] + beta[6]*pr.spoc1[i]+beta[7]*spanish2[i]
}
#
for(k in 1:7){
beta[k] ~ dnorm(0.0,1.0E-6)
}

DATA
→←

INITS
list(beta=c(0,0,0,0,0,0))

      mean   sd  MC_error  val2.5pc  median  val97.5pc start sample
beta[1]  1.429  0.106  9.653E-4  1.222  1.43  1.635  1001  20000
beta[2] -0.06401  0.01359  1.202E-4 -0.09096 -0.06391 -0.03748  1001  20000
beta[3] -0.5647  0.1884  0.001422 -0.9363 -0.564 -0.1989  1001  20000
beta[4] -0.1886  0.08691  6.788E-4 -0.3615 -0.1879 -0.01858  1001  20000
beta[5] -0.7163  0.1145  9.81E-4 -0.9408 -0.7161 -0.4942  1001  20000
beta[6] -0.4514  0.08761  6.941E-4 -0.6213 -0.4515 -0.2797  1001  20000
beta[7] -0.6057  0.08288  6.98E-4 -0.7668 -0.6061 -0.4453  1001  20000
deviance 4342.0  3.761  0.03185  4337.0  4342.0  4352.0  1001  20000
```

Summary