Workshop on Bayesian Modeling Using WinBUGS

Sessions 8–9 (A): : Normal Linear Models

Workshop web-page: `http://www.math.ntua.gr/~fouskakis/Workshop/home.html`

Workshop on Bayesian Modeling Using WinBUGS

Sessions 8–9 (A): : Normal Linear Models

Ioannis Ntzoufras

Department of Statistics,
Athens University of Economics and Business,
Athens, Greece;

e-mail: `ntzoufras@aueb.gr`.

Workshop web-page: `http://www.math.ntua.gr/~fouskakis/Workshop/home.html`

This presentation is based on Chapter 5 of Ntzoufras (2009): *Bayesian Modeling Using WinBUGS*, Wiley.

Synopsis

1. General modeling principles
2. Model specification for normal regression models
3. Using vectors and multivariate priors
4. ANOVA models (one way and parametrizations)

5.1 General modeling principles

Statistical Model

- DESCRIBES parsimoniously real life problems observed under uncertainty.
- is a collection of probabilistic statements (and equations) that describe and interpret present behavior or predict future performance.

Three important components

1. Response variable (or variables) Y,
2. Explanatory variables $X_1, X_2, \ldots, X_p$,
3. Linking mechanism between the two sets of variables.

The response variable Y

$\Rightarrow$ is the main study variable,

$\Rightarrow$ it represents the stochastic part of the model.

`stochastic` refers to $\Rightarrow$ random variables whose outcome is uncertain before it is observed.

We are frequently interested in

1. describing the mechanism leading to the appearance of a certain outcome of Y
2. predicting a future outcome of Y.

The Stochastic Component

$$Y|X_1, X_2, \ldots, X_p \sim \mathcal{D}(\boldsymbol{\theta})$$

where $\mathcal{D}(\boldsymbol{\theta})$ is a distribution with parameter vector $\boldsymbol{\theta}$. For normal regression models:

$$Y|X_1, X_2, \ldots, X_p \sim N(\mu, \sigma^2),$$

where $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

- One response variable $\Rightarrow$ *univariate* model.
- More than one response variables $\Rightarrow$ *multivariate* model.

Explanatory Variablaes

- Explanatory variables $X_1, \ldots, X_p \Rightarrow$ all variables that potentially influence Y.
- The main focus in such models is to infer concerning
  1. the significance,
  2. the type (negative or positive), and
  3. the magnitude of the effect of each X_i on Y is the main focus in such models.

The Stochastic Component (again)

- (Usually) X_i are considered as fixed, nonstochastic components, that is, deterministic nodes in WinBUGS .
- This is the reason of defining the distribution of Y conditional on the observed explanatory variables

$$Y|X_1, \ldots, X_p \sim \mathcal{D}\Big(\boldsymbol{\theta}(\boldsymbol{\beta}, \boldsymbol{\phi}, X_1, \ldots, X_p)\Big) \ .$$

- Parameter vector $\boldsymbol{\theta}$ is
  - refers to the parameters of the distribution of Y
  - it is expressed as a function of the explanatory variables and a new alternative set of parameters $(\boldsymbol{\beta}, \boldsymbol{\phi})$ that substitutes the original ones in terms of estimation and inference.
- $\boldsymbol{\beta}$ summarizes the association between the response and the explanatory variables,

- ϕ refers to other characteristics of the distribution such as the variance or the shape.
- (Usually) The mean of the response model is associated with the response variables, but in more complicated models, the variance or other moment functions can also be estimated via the explanatory variables.

Linking the Response with the Covariates

- Use a function to connect the parameters describing Y (e.g. mean) with covariates.
- Simpler case $\Rightarrow$ express the mean of Y as a function of the linear combination of the explanatory variables:

$$E(Y|X_1, X_2, \ldots, X_p) = \mu(\boldsymbol{\beta}, X_1, \ldots, X_p) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} \beta_j X_j\right).$$

  Here, the linear combination of X_is
  - is used to predict the $E(Y)$
  - It is called the *linear predictor* η of the model
- Terminology and principles were originally introduced in generalized linear models (McCullagh and Nelder, 1989), but they can be adopted for a wide range of models. Within this context, $g(\mu)$ is referred as the *link function*.

Model Structure

General

1. Stochastic Component: $Y \sim D(\boldsymbol{\theta})$
2. Deterministic component: Covariates $X_1, \ldots, X_p$.
3. Link function: $\boldsymbol{\theta} = g^{-1}(X_1, \ldots, X_p)$.

Normal linear regression model

1. Stochastic Component: $Y \sim N(\mu, \sigma^2)$
2. Deterministic component: Covariates $X_1, \ldots, X_p$ and linear predictor $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$.
3. Link function: $\mu = \eta \Leftrightarrow \mu = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$.

The prior distribution

- To complete the Bayesian formulation, a prior distribution $f(\boldsymbol{\beta}, \boldsymbol{\phi})$ for the parameters $(\boldsymbol{\beta}, \boldsymbol{\phi})$ remains to be specified (will be discussed later)
- Expresses prior information or knowledge about model parameters
- Usually no prior information is available and hence "vague" or "flat" non-informative priors can be used.

Random Covariates?

- Usually defined as deterministic (fixed) quantities.
- In practice are frequently random.
- The model can be extended by considering random X_i with additional parameters under estimation; see Ryan (1997, pp. 34–35) for a discussion.
- This can be specified within the Bayesian framework using additional hierarchical levels in our model (see Chapter 9 of my book for details).

5.2 Model specification in normal regression models

- The most popular models in statistical science.
- Based on the initial work of Sir Francis Galton in the late years of the 19th century (Stanton, 2001).
- Y is considered a continuous random variable defined in the whole set of real numbers
- We assume that Y follows the normal distribution with mean μ and variance σ^2.

Regression Model Structure for n observations

This model structure holds for every observation $i = 1, 2, \ldots, n$ and it is denoted with a subscript i:

$$\begin{aligned} Y_i &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 X_{i1} + \cdots + \beta X_{ip} \\ &\quad \text{for} \;\; i = 1, \ldots, n \end{aligned} \tag{1}$$

- $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_n, \sigma^2)$ parameters describing Y (of length $n + 1$)
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ describes the linear effect of X_is on Y (of length $p + 1$)
- $\phi = \sigma^2$ describes the variance of the around the predicted mean.
- $\tau = 1/\sigma^2$ described the precision (accuracy) of the predictions based on the expected mean of the model.

Alternative Expression

Frequently, the following alternative representation of the regression model is adopted

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2), \text{for} \;\; i = 1, \ldots, n.$$

- Nice interpretation (Y = function of the X_is and a random normal error with variance σ^2)
- Expression (1) is more general and follows the model building principles described in Section 5.1 of this presentation.

Notation simplification

- The following notation is also used in WinBUGS
- Remove the condition on the explanatory variables.
  - Denote $Y|X_1, \ldots, X_p$ simply by Y
  - $E(Y|X_1, \ldots, X_2)$ by $E(Y)$ or μ.
- In the normal distribution (and the regression model), we work with the precision parameter $\tau = 1/\sigma^2$

5.2.1 Specifying the likelihood

Let us observe a sample of size n with

- response values $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ and
- $x_{i1}, \ldots, x_{ip}$, the values of the explanatory variables $X_1, \ldots, X_p$ for individuals $i = 1, \ldots, n$.

The model will be now written in WinBUGS:

for $i = 1, \ldots, n$

$Y_i \sim N(\mu_i, \tau^{-1})$

$\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$

$\sigma^2 = 1/\tau$

$\sigma = \sqrt{\sigma^2}$

```
for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- beta0 + beta1*x1[i] + ... +
        betap*xp[i]
}
s2<-1/tau
s <-sqrt(s2)
```

Linear predictor - notation

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

```
mu[i] <- beta0 + beta1*x1[i] + ... +
    betap*xp[i]
```

- Explanatory variables ⇒ vector nodes with names `x1, ..., xp` of length n
- $\boldsymbol{\mu}$ ⇒ vector node with names of length n
- All β_j are defined separately as single scalar nodes
- Each parameter must be monitored separately in the `sample monitor` tool of WinBUGS

Variance specification - deterministic nodes

$$\sigma^2 = 1/\tau$$
$$\sigma = \sqrt{\sigma^2}$$

```
s2<-1/tau
s <-sqrt(s2)
```

- The last two commands are used to deterministically specify the relation between the variance, the standard deviation, and the precision parameter $\tau = \sigma^{-2}$ used by the normal distribution in WinBUGS.

5.2.2 Specifying a simple independent prior distribution

In normal regression models, the simplest approach is to assume that a priori all parameters are independent having the structure

$$\begin{aligned} f(\boldsymbol{\beta}, \tau) &= \prod_{j=0}^{p} f(\beta_j) f(\tau), \\ \beta_j &\sim N(\mu_{\beta_j}, c_j^2) \text{ for } j = 0, \ldots, p \text{ and} \\ \tau &\sim \text{gamma}(a, b) \ . \end{aligned} \tag{2}$$

- The variance σ^2 is substituted by the corresponding precision parameter τ in order to make it compatible to the WinBUGS notation.
- The gamma prior of $\tau \Rightarrow E(\tau) = \frac{a}{b}$ and $\text{Var}(\tau) = \frac{a}{b^2}$.
- Prior for $\sigma^2 \Rightarrow$ inverse gamma prior distribution with

$$E(\sigma^2) = \frac{b}{a-1} \quad \text{and} \quad \text{Var}(\sigma^2) = \frac{b^2}{(a-1)^2(a-2)}.$$

Prior when no information is available

- A usual choice for the prior mean is zero ($\mu_{\beta_j} = 0$).
  - Centers our prior beliefs around zero $\Rightarrow$ corresponds to the assumption of no effect of X_j on Y.
  - We express our prior doubts about the effect of X_j on Y, prompting Spiegelhalter et al. (2004, pp. 90, 158–160) to call this a "sceptical" prior.
- The prior variance c_j^2 of the effect β_j is set equal to a large value (e.g., 10^4) to represent high uncertainty or prior ignorance.
- Similarly, for τ we use equal low prior parameter values, setting in this way
  - Prior mean equal to one: $E(\tau) = 1$
  - Prior variance large: $E(\tau) =$ large.
  - For example, $a = b = 0.01$ results in $E(\tau) = 1$ and $V(\tau) = 100$.
  - This approach is adopted in all illustrations of the WinBUGS manual and example volumes.

Within WinBUGS , the prior setup is specified with the following syntax

$\beta_0 \sim N(0, 10^4)$
$\beta_1 \sim N(0, 10^4)$
.....
$\beta_p \sim N(0, 10^4)$
$\tau \sim \text{gamma}(0.01, 0.01)$

```
beta0 ~ dnorm( 0.0, 1.0E-4 )
beta1 ~ dnorm( 0.0, 1.0E-4 )
.........................
betap ~ dnorm( 0.0, 1.0E-4 )
tau   ~ dgamma( 0.01, 0.01 )
```

- Value `1.0E-4` is the scientific notation for $1.0 \times 10^{-4} = 0.001$, (prior precision of each β_j). It corresponds to prior variance equal to 10^4.
- The definition above can be considerably simplified by using vectors instead of single nodes; see Section 5.3.3 for details.

5.2.3 Interpretation of the regression coefficients

- Each β_j refers to the effect of X_j on the expectation of Y adjusted for the remaining covariates.
- Inference of model parameters can be divided into three basic stages:
  1. Is the effect of X_j important for the prediction or description of Y?
  2. What is the association between Y and X_j (positive, negative, or other)?
  3. What is the magnitude of the effect of X_j on Y ?

Importance of the Effect

- Is the posterior distribution of β_j is scattered around zero (or not)? Posterior distributions away from zero $\Rightarrow$ important effect of X_j on Y.
- Such analysis offers a first rough tool for tracing important variables.
- Within this analysis, we can calculate the posterior probability:

$$\pi_0 = \min\left\{ f(\beta_j < 0|\boldsymbol{y}), f(\beta_j > 0|\boldsymbol{y}) \right\}.$$

When zero lies at the center of the posterior distribution $\Rightarrow \pi_0 \to \frac{1}{2} \Rightarrow$ no clear positive or negative effect of X_j on Y.

In WinBUGS we calculate $f(\beta_j > 0|\boldsymbol{y})$ using the syntax

```
p.betaj <- step( betaj )
```

Obtaining the posterior mean via the sample monitor tool provides us the estimate of the posterior probability $f(\beta_j > 0|\boldsymbol{y})$.

Importance of the Effect & Model Comparison

- Deviance information criterion (DIC) (Spiegelhalter et al., 2002)
  - Available in WinBUGS
  - Compares two models (e.g. with and without a covariate)
  - Similar to AIC
  - Very popular nowadays because it can be easily implemented
- Formal Bayesian hypothesis testing is based on **posterior model odds** and **Bayes factors**
  - Difficult to describe in the level
  - You can look at Chapter 11 of my book for examples (and a older publication in JSS)

Type of Association

- Is the relationship is positive or negative?
- Examine the signs of the posterior summaries of central and relative location (e.g., mean, median, 2.5% and 97.5% percentiles).
- If all of them are positive or negative, then the corresponding association can be concluded.
- *Positive association* $\Rightarrow$ changes of X_j cause changes of the same direction for Y
- *Negative association* $\Rightarrow$ changes of X_j cause changes of the opposite direction for Y.
- Within this analysis, we can also use $\pi_0 = \min\left\{f(\beta_j < 0|\boldsymbol{y}), f(\beta_j > 0|\boldsymbol{y})\right\}$. When π_0 is low (e.g., $\leq 2.5\%$, 1%, or 0.5%) $\Rightarrow$ positive or negative association depending on the sign of the posterior location summaries.

Magnitude of the Effect
Interpretation of β_j for $j \geq 1$

- β_j captures the magnitude of the effect of X_j on Y.
- We can obtain this by considering the difference of the expected values for $X_j = x + 1$ and $X_j = x$ without changing the remaining covariates.

$$\begin{aligned}
\Delta\mu_{X_j} &= \mu(\boldsymbol{\beta}, X_1, \ldots, X_{j-1}, \mathbf{X_j = x + 1}, X_{j+1}, \ldots, X_p) \\
&\quad -\mu(\boldsymbol{\beta}, X_1, \ldots, X_{j-1}, \mathbf{X_j = x}\quad\;\;, X_{j+1}, \ldots, X_p) \\
&= \beta_0 + \beta_1 X_1 + \cdots + \beta_{j-1} X_{j-1} + \beta_j (x+1) + \beta_{j+1} X_{j+1} + \cdots + \beta_p X_p \\
&\quad -\beta_0 - \beta_1 X_1 - \cdots - \beta_{j-1} X_{j-1} - \beta_j x \qquad\quad - \beta_{j+1} X_{j+1} - \cdots - \beta_p X_p \\
&= \beta_j \; .
\end{aligned}$$

- Posterior mean or median of β_j refers to the corresponding posterior measures of the expected change of Y.

- An increase of one unit of X_j, *given that the remaining covariates will remain stable*, induces an a posteriori average change on the expectation of Y equal to the posterior mean of β_j; see Table 5.1 for further details.

Table 5.1: Summary interpretation table for regression coefficients β_j

INTERPRETATION OF MODEL COEFFICIENTS β_j $(j = 1, \ldots, p)$

$$\Delta\mu_{X_j} = \beta_j$$

- $\Delta\mu_{X_j}$ denotes the expected difference of Y if X_j increases by one unit and the rest of the covariates remain the same.
- If $\beta_j = 0 \Rightarrow$ no effect on Y.
- If $\beta_j < 0 \Rightarrow$ negative effect on Y (Y is expected to decrease when X_j increases and vice versa).
- If $\beta_j > 0 \Rightarrow$ positive effect on Y (Y is expected to increase when of X_j increases and decrease when X_j decreases).
- β_j is the expected change (increase or decrease) when X_j increases by one unit and the rest of the covariates remain unchanged.

Interpretation of β_0

β_0 is the expected value of Y when all $X_j = 0$

- Such combination ($X_j = 0$ for all $j = 1, \ldots, p$) drequently lies outside the range of the observed covariate values.
- Interpretation of β_0 is not reliable since we predict Y for values of X_j that have not been observed.
- Direct interpretation of β_0 may not lead to realistic and sensible interpretation.

Interpretation of β_0 - Expected Y for a Typical Subject

Alternatively we may center all X_j by subtracting their sample mean:

- β_0^c represents the expected value of Y when $X_j = \overline{X}_j \Rightarrow \beta_0^c$ is the expected response Y for an "average" or "typical" subject according to our sample.

In WinBUGS:

```
typical.y<- beta0 + beta1 * mean(x1[]) + ... + betap * mean(xp[])
```

without changing the parametrization of the original model.

The approach described above can also be used to calculate the expected values of Y for any combination of values of X_j.

Interpretation of τ or σ^2

Parameter τ (and the variance σ^2) indicates the precision of the model.

- If τ is high (σ^2 low), then the model can accurately predict (or describe) the expected values of Y.
- According to the properties of the normal distribution
  - $\mu_i \pm \sigma$ includes 66% of the values of Y
  - $\mu_i \pm 2\sigma$ includes 95% of the values of Y
  - $\mu_i \pm 3\sigma$ includes 99% of the values of Y

Interpretation of τ or σ^2

- We can rescale this τ or σ^2 using the sample variance of Y, namely, s_Y^2:

$$R_B^2 = 1 - \frac{\tau^{-1}}{s_Y^2} = 1 - \frac{\sigma^2}{s_Y^2}.$$

- $R_B^2 \Rightarrow$ proportional reduction of uncertainty concerning the response variable Y achieved by incorporating X_js in the model.
- Is the Bayesian analog of R^2.
- WinBUGS syntax

```
sy2<- pow( sd(y[]) , 2)
R2B <- 1 - s2/sy2
```

or

```
R2B <- 1 - 1/(tau*sy2)
```

using the precision parameter τ

5.2.4 A regression example using WinBUGS

Example 5.1. Soft drink delivery times.

- *Example deals with the quality of the delivery system network of a soft drink company; see example 4.1 in Montgomery and Peck (1992).*
- *We are interested in estimation of the required time needed by each employee to refill an automatic vending machine.*
- *Response variable ⇒ total service time (measured in minutes) of each machine*
- *Covariates*
  1. *the number of cases of stocked products and*
  2. *the distance walked by the employee (measured in feet).*
- *25 observations*
- *This dataset is reproduced in the book's Website with permission of John Wiley and Sons, Inc.*

Setting up the data and the model code

- Model code according to Sections 5.2.1 and 5.2.2
- Data can be defined in either a rectangular or a list format.
- Rectangular format of the data are provided in Table 5.2
- Full model code of the example, including the list data format and the initial values, is given in Table 5.3.
- All three variables used in the model (time, cases, distance) are defined as separate vectors in the list data format
- In the initial values, each parameter τ, β_0, β_1, and β_2 was initialized separately.

Table 5.2: Rectangular data WinBUGS format of "soft drink delivery times" example

```
time[]  cases[] distance[]
 16.68   7 560
 11.5    3 220
 12.03   3 340
 14.88   4 80
 13.75   6 150
 18.11   7 330
  8      2 110
 17.83   7 210
 79.24  30 1460
 21.5    5 605
 40.33  16 688
 21     10 215
 13.5    4 255
 19.75   6 462
 24      9 448
 29     10 776
 15.35   6 200
 19      7 132
 9.5     3 36
 35.1   17 770
 17.9   10 140
 52.32  26 810
 18.75   9 450
 19.83   8 635
 10.75   4 150
END
```

← Empty line

Table 5.3: Full model code for "soft drink delivery times" example

```
model{
    # model's likelihood
    for (i in 1:n){
         time[i] ~ dnorm( mu[i], tau ) # stochastic componenent
         # link and linear predictor
         mu[i] <- beta0 + beta1 * cases[i] + beta2 * distance[i]
    }
    # prior distributions
    tau ~ dgamma( 0.01, 0.01 )
    beta0 ~ dnorm( 0.0, 1.0E-4)
    beta1 ~ dnorm( 0.0, 1.0E-4)
    beta2 ~ dnorm( 0.0, 1.0E-4)
    # definition of sigma
    s2<-1/tau
    s <-sqrt(s2)
    # calculation of the sample variance
    for (i in 1:n){ c.time[i]<-time[i]-mean(time[]) }
    sy2 <- inprod( c.time[], c.time[] )/(n-1)
```

```
    # calculation of Bayesian version R squared
    R2B <- 1 - s2/sy2
    # Expected y for a typical delivery time
    typical.y <- beta0 + beta1 * mean(cases[]) + beta2 * mean(distance
        [])
}
INITS
list( tau=1, beta0=1, beta1=0, beta2=0 )

DATA (LIST)
list( n=25,
      time = c(16.68, 11.5, 12.03, 14.88, 13.75, 18.11,  8, 17.83,
               79.24, 21.5, 40.33, 21, 13.5, 19.75, 24, 29, 15.35,
               19, 9.5, 35.1, 17.9, 52.32, 18.75, 19.83, 10.75),
      distance = c(560, 220, 340, 80, 150, 330, 110, 210, 1460,
                   605, 688, 215, 255, 462, 448, 776, 200, 132,
                   36, 770, 140, 810, 450, 635, 150),
      cases = c( 7, 3, 3, 4, 6, 7, 2, 7, 30, 5, 16, 10, 4, 6, 9,
                 10, 6, 7, 3, 17, 10, 26, 9, 8, 4) )
```


Results

- 3000 iterations and discarding the initial 1000 ones
- Posterior summaries are provided in Table 5.4
- Posterior densities are provided in Figure 5.1
- Descriptive analysis of the posterior distribution of R^2_B indicates a considerable improvement of the precision (posterior mean equal to 0.95) in the prediction of delivery times when including in the model covariates cases and distance.

Table 5.4: WinBUGS posterior summaries for Example 5.1 after 2000 iterations and additional discarded 1000 burnin iterations

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
R2B	0.9511	0.01743	5.118E-4	0.9064	0.9548	0.9734	1001	2000
beta0	2.356	1.188	0.03076	-0.03996	2.372	4.635	1001	2000
beta1	1.61	0.1806	0.003737	1.272	1.609	1.968	1001	2000
beta2	0.01447	0.003812	8.476E-5	0.006872	0.01446	0.02211	1001	2000
p.beta0	0.974	0.1591	0.004037	0.0	1.0	1.0	1001	2000
p.beta1	1.0	0.0	2.236E-12	1.0	1.0	1.0	1001	2000
p.beta2	1.0	0.0	2.236E-12	1.0	1.0	1.0	1001	2000
s	3.386	0.5695	0.0168	2.531	3.302	4.749	1001	2000
typical.y	22.38	0.683	0.01701	21.09	22.37	23.78	1001	2000

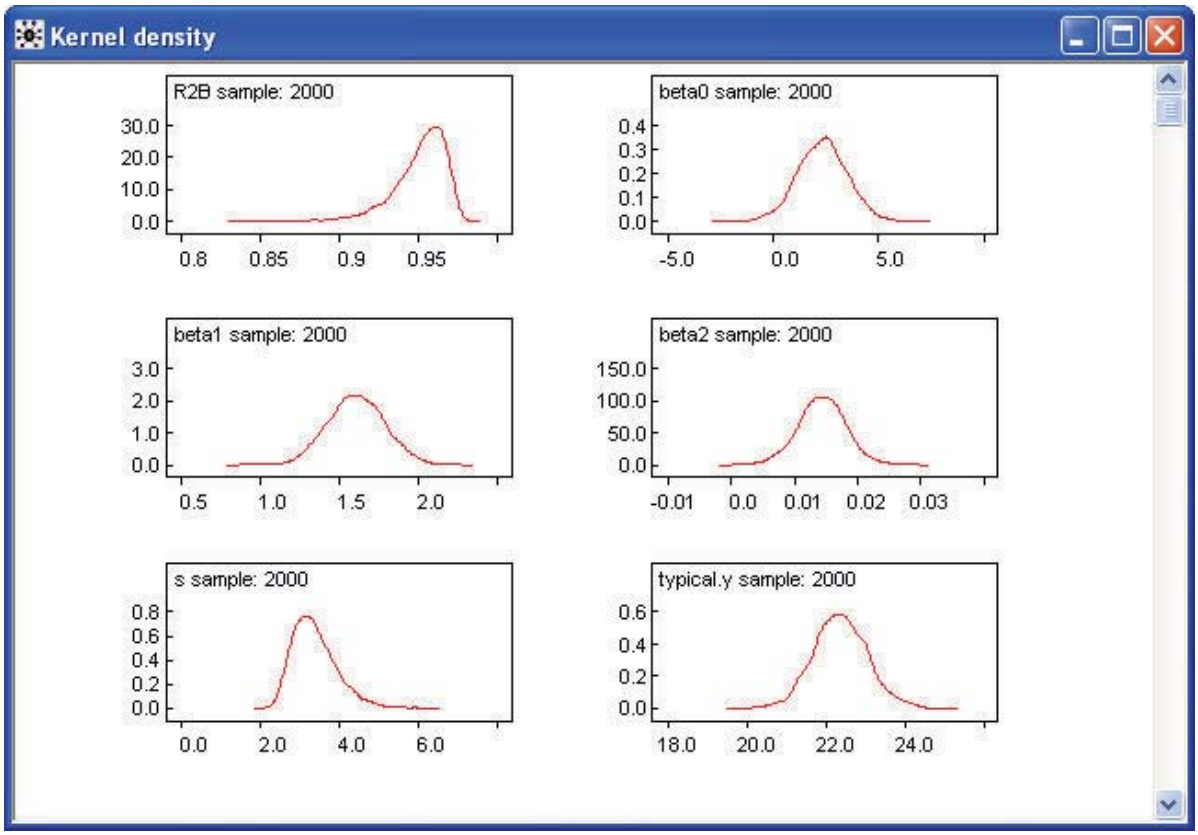


Figure 5.1: Posterior densities of model parameters for Example 5.1

Results - Parameter Interpretation

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
s	3.386	0.5695	0.0168	2.531	3.302	4.749	1001	2000

- Posterior distribution of σ: with the current model we can predict the expected delivery time with an a-posteriori expected error of 3.4 minutes

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	2.356	1.188	0.03076	-0.03996	2.372	4.635	1001	2000
beta1	1.61	0.1806	0.003737	1.272	1.609	1.968	1001	2000
beta2	0.01447	0.003812	8.476E-5	0.006872	0.01446	0.02211	1001	2000
p.beta0	0.974	0.1591	0.004037	0.0	1.0	1.0	1001	2000
p.beta1	1.0	0.0	2.236E-12	1.0	1.0	1.0	1001	2000
p.beta2	1.0	0.0	2.236E-12	1.0	1.0	1.0	1001	2000

- Regression line based on posterior means:

$$\text{Expected time } = 2.36 + 1.6 \times \text{ cases } + 0.015 \times \text{ distance.}$$

- Regression line based on posterior medians: similar
- The effect of both explanatory variables (cases and distance) have an important contribution to the prediction of delivery time.
- All summary statistics and the posterior densities indicate that zero is far away from the posterior distribution with posterior probability of having positive association between each X_j and Y equal to one.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta1	1.61	0.1806	0.003737	1.272	1.609	1.968	1001	2000
beta2	0.01447	0.003812	8.476E-5	0.006872	0.01446	0.02211	1001	2000

- Each additional case stocked by the employee ⇒ +1.6 minutes (+96 secs) expected delivery time
- This increase lies between 1.3 and 2.0 minutes (76 and 118 seconds) with probability 95%.
- Increase of one foot walking distance ⇒ +0.87 seconds of posterior mean of expected delivery time
- Additional 100 feet of walking ⇒ +1.5 minute of posterior mean of expected delivery time (0.7–2.2 minutes with probability 95%)
- Additional 100 meters of walking distance ⇒ +4.7 minutes of posterior mean of expected delivery time

  (one foot is equal to 0.3048 m, resulting in an increase of the expected delivery time by $100/0.3048 * 0.01447 = 4.747$ minutes).

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	2.356	1.188	0.03076	-0.03996	2.372	4.635	1001	2000
typical.y	22.38	0.683	0.01701	21.09	22.37	23.78	1001	2000

- Parameter β_0 has no sensible interpretation in this example since the zero value is nonsense for both explanatory variables

  (the delivery employee will always have to stock some cases of products in the machine and walk at least a small distance to reach the delivery location).
- For this reason, no interpretation of this parameter is attempted.
- Zero lies at the left tail of the posterior distribution within the range of the 95% posterior interval.
- The posterior probability of positive β_0 is equal to 97.4%.
- Hence we may focus on the predicted value for a typical or delivery route.
- According to the posterior summaries of `typical.y`, a typical delivery route will take 22.4 minutes on average and will range from 21.1 to 23.8 minutes with probability 95%.

5.3 Using vectors and multivariate priors in normal regression models

5.3.1 Defining the model using matrices

- The regression model can be defined using vectors and matrices instead of scalar nodes.
- Substitute the mean code line with the following syntax:

```
mu[i] <- beta0 + beta[1] * x[i,1] + ... + beta[p]*x[i,p]
```

  - Matrix $\boldsymbol{x}$ (denoted by `x[,]` in WinBUGS) is a $n \times p$ matrix.
  - Each column $\boldsymbol{x}_j$ of $\boldsymbol{x}$ corresponds to each explanatory variable X_j,
  - Each row $\boldsymbol{x}_{(i)}$ corresponds to the explanatory variable values of the ith subject of the sample.

- Data definition:
  - In rectangular data format - each column of the matrix can be defined using the with header `x[,1] ... x[,p]`.
  - In list format - `x` is defined as an array with dimensions n and p.
- When the number of variables p is large, we can use the command

$$\texttt{inprod(b[], x[i,])} \Rightarrow \sum_{j=1}^{p} \beta_j x_{ij}$$

  Hence, the mean can be defined by the syntax

```
mu[i] <- beta0 + inprod( b[], x[i,])
```

  where `b[]` is vector $\boldsymbol{\beta}_{\{\setminus 0\}} = (\boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_p)$.

- Usually the constant term is included in a matrix $\boldsymbol{X}$ by setting

$$\boldsymbol{X} = [\mathbf{1}_n, \boldsymbol{x}]$$

- $\boldsymbol{X}$ is of dimension $n \times (p+1)$ and is called the *data matrix* (or design matrix).
- $\mathbf{1}_n$ is a vector of length n with all elements equal. Corresponds to the "constant" term.
- The linear predictor can be now written as $\mu_i = \boldsymbol{X}_{(i)}\boldsymbol{\beta}$ where $\boldsymbol{X}_{(i)}$ is the ith row of $\boldsymbol{X}$.
- This expression can be coded in WinBUGS (within the likelihood for loop) by

```
x[i,1]<-1
mu[i] <- inprod( X[i,], beta[] )
```

With this approach we can monitor all regression coefficients simultaneously by simply considering the vector node `beta`.

- This expression can be coded in WinBUGS (within the likelihood for loop) by

```
x[i,1]<-1
mu[i] <- inprod( X[i,], beta[] )
```

- `X[,]` $n \times (p+1)$ matrix with all components of the first column `x[,1]` equal to one.
- `beta[1]` represents the constant coefficient β_0,
- `X[,j]` refers to the vector of values of X_{j-1}
- `beta[j]` corresponds to coefficient β_{j-1} for $j = 1, \ldots, p+1$
- Rectangular data format: use header names `X[,2] ... X[,p+1]`.
- List data format: data are as matrix defined in the usual manner with the elements of the 1st column equal to
  - `NA` (specification within the code)
  - one (specification within the data)

5.3.2 Prior distributions for normal regression models

The Conjugate Normal–Inverse Gamma Prior

- Conjugate analysis for the normal regression model has been presented in Section 1.5.5 of my book; also see in Ntzoufras (2010).
- The conjugate normal–inverse gamma prior is considered if

$$\boldsymbol{\beta}|\sigma^2 \sim N_P(\boldsymbol{\mu}_\beta, c^2\boldsymbol{V}\sigma^2) \quad \text{and} \quad \sigma^2 \sim \text{IG}(a,b) \ , \tag{3}$$

  - P=$p+1$
  - c^2 is a parameter controlling the overall magnitude of the prior variance.

Conjugate Prior - Zellner's g-Prior

- Special case: Zellner (1986) g-prior for $\boldsymbol{V} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$.
- Parameter $c^2 \Rightarrow g$ in Zellner's original publication.
- Default choice: $c^2 = n$
  - Usually adopted when no information is available
  - It has an interpretation of adding prior information equivalent to one data point; (Kass and Wasserman, 1995; Fouskakis et al., 2009).
- This prior has been widely used because
  - it considerably simplifies posterior computations
  - reduces the number of prior variance parameters that remain to be specified down to one.

Conjugate Prior - Independence between β_js

- When no prior information is available, we may simplify the prior by considering independent normal distributions by setting
  - $\boldsymbol{V} = c^2 \boldsymbol{I}_P$
  - c^2 large to express prior ignorance (e.g., $c = 100$).
- Hence we can simply rewrite the prior as

$$\beta_j|\sigma^2 \quad \sim \quad N(\mu_{\beta_j}, c^2\sigma^2) \text{ for } j = 0, 1, \ldots, p, \tag{4}$$

  where μ_{β_j} are the components of the prior mean vector $\boldsymbol{\mu}_\beta$.
- This setup is also related to the *ridge regression*.

Conjugate Prior and Variable Selection

- The conjugate prior setup is very convenient for implementing Bayesian variable selection (Raftery et al., 1997).
- Zellner's g-priors were widely used within this context since they allow us for a sensible default choice of prior distributions; see
  - Fernandez et al. (2000) for comparison between different values of c^2
  - Liang et al. (2008) for discussion and extensions concerning the g-priors.

Conditional Conjugate Priors

- Simpler prior setup

$$\boldsymbol{\beta} \sim N_P(\boldsymbol{\mu}_\beta, c^2\boldsymbol{V}) \quad \text{and} \quad \sigma^2 \sim \text{IG}(a, b)$$

  i.e. $\boldsymbol{\beta}$ and σ^2 are a priori independent

- Even simpler: if we a priori assume that all β_j and σ^2 are independent. This is usually adopted when no information is available.
- In this case the prior is not conjugate and hence MCMC methods need to be implemented in order to estimate the posterior distribution.
- Nevertheless, this is conditionally conjugate, resulting in conditional posterior distributions for $\boldsymbol{\beta}$ and τ that can be calculated analytically, allowing us to construct an efficient Gibbs sampler (it is done internally in WinBUGS)

Other Priors

Finally, other type of prior distributions for $\boldsymbol{\beta}$ have been proposed in the related literature. For example,

- the Student t distribution
- the Cauchy distribution
- the double exponential (Bayesian LASSO)

can be used instead of the normal prior, but obvious differences are seldom observed when no prior information is available.

5.3.3 Multivariate normal priors in WinBUGS

Syntax for Independent Normal Priors Using Vectors

- Independence prior (2) can be defined in WinBUGS by specifying the priors for β_j terms within a loop:

for $j = 1, \ldots, P$ $\beta_j \sim N(0, 10000)$	`for (j in 1:P){` `  beta[j] ~ dnorm( 0.0, 1.0E-4 ) }`

- Conjugate prior (4):

$\xi = 1/(c^2\sigma^2) = \tau/c^2$
for $j = 1, \ldots, P$
$\beta_j \sim N(\beta_{0j}, 1/\xi)$
	`xi <- tau/c2`
	`for (j in 1:P){`
	`  beta[j] ~ dnorm( beta0[j], xi ) }`

Syntax for Multivariate Conjugate Prior

- Multivariate Normal–Inverse Gamma prior (3).
- Specify (in WinBUGS) the precision matrix $\boldsymbol{T}$ of dimension $P \times P$
- The elements of $\boldsymbol{T}$ are given by

$$T_{lj} = \frac{\tau}{c^2}[\boldsymbol{V}^{-1}]_{lj} \tag{5}$$

for $l, j \in \{1, 2, \ldots, P\}$,

- $\boldsymbol{V}^{-1}$ is the inverse of $\boldsymbol{V}$
- $[\boldsymbol{V}^{-1}]_{lj}$ is the lth row and jth column element of $\boldsymbol{V}^{-1}$

In WinBUGS

1. Calculate $\boldsymbol{V}^{-1}$ using the syntax

```
inverse.V[1:P,1:P] <- inverse(V[,])
```

2. Calculate the elements of the prior precision matrix using (5) within a double `for` loop:

```
for(l in 1:P){
   for (j in 1:P){ prior.T[l,j] <- inverse.V[l,j] * tau /c2 }}
```

3. Complete the prior specification using the following syntax

$\boldsymbol{\beta} \sim N_P(\boldsymbol{\mu}_\beta, \boldsymbol{T}^{-1})$
$\tau \sim \text{gamma}(1/100, 1/100)$
$\sigma^2 = 1/\tau$

```
beta[1:P] ~ dmnorm( mu.beta[], prior.T[,] )
tau     ~ dgamma( 0.01, 0.01 )
s2      <- 1/tau
```

– Prior values for $\boldsymbol{V}$, $\boldsymbol{\mu}_\beta$, and c^2 can be specified either within the data list or directly within the WinBUGS model code.

Specification of the Prior Variance-Covariance Matrix
Independent priors as a special case

The independence prior (4) can be defined within the WinBUGS syntax by specifying

$c^2 = 100$
$\boldsymbol{\mu}_\beta = \mathbf{0}_P$
for l=1, . . . , P and j=1, . . . , P
$V_{lj} = I(l = j)$
(i.e. $\boldsymbol{V} = \boldsymbol{I}_P$)

```
c2 <- 100
for (j in 1:P){ mu.beta[j] <- 0.0 }
for (l in 1:P){ for (j in 1:P){
    V[l,j] <- equals(l,j)
}}
```

Node `V[l,j]` takes the value one if $l = j$ (hence we have a diagonal element) and zero otherwise.

Specification of the Prior Variance-Covariance Matrix
Zellner's g-prior

Zellner's g-prior can be specified using the syntax

for l=1, . . . , P and j=1, . . . , P
$[V_{lj}]^{-1} = \boldsymbol{X}_l^T \boldsymbol{X}_j = \sum_{i=1}^{n} x_{il} x_{ij}$
(i.e. $\boldsymbol{V} = \boldsymbol{X}^T \boldsymbol{X}$)

```
for (l in 1:P){ for (j in 1:P){
    inverse.V[l,j]<-inprod( X[,l], X[,j] )
}}
```

$[\boldsymbol{V}^{-1}]_{lj}$ are the elements of matrix $(\boldsymbol{X}^T \boldsymbol{X})$ used in Zellner's g-prior.

5.3.4 Continuation of Example 5.1

- Here we rerun the same model using Zellner's g-prior.
- The basic code of the model is given in Table 5.5, in which matrix $\boldsymbol{V} = (\boldsymbol{X}^T \boldsymbol{X})$ is defined within the model code.
- The first column `X[,1]` of matrix `X` is defined within the model code, and the remaining columns are defined within the data part.
- In Table 5.5, the first and last rows of the data are also given (in a rectangular data format) while the remainder have been denoted by dotted lines to save space.
- Additionally, the sample size `n` and the number of parameters `P` involved in the linear predictor are specified separately in a list format (hence these two types of data must be loaded separately).

Table 5.5: WinBUGS code for Example 5.1 using Zellner's g-prior and parameter vectors (Data are compressed to conserve space)

```
model{
    # definition of prior parameters
    c2 <- 10000
    for (j in 1:P){ mu.beta[j] <- 0.0 }  # prior mean
    # calculation of xtx
    for (i in 1:P){ for (j in 1:P){
       inverse.V[l,j] <- inprod( X[,l] , X[,j] )    }}
    # calculation of the elements of prior precision matrix
    for(l in 1:P){ for (j in 1:P){
      prior.T[l,j] <- inverse.V[l,j] * tau /c2      }}
    # model's likelihood
    for (i in 1:n){
         X[i,1] <- 1.0
         # specifying the constant term in the first column
         time[i] ~ dnorm( mu[i], tau ) # stochastic componenent
         # link and linear predictor
         mu[i] <- inprod( beta[], X[i,] )
    }
```



```
    # prior distributions
    # multivariate prior for the beta vector
    beta[1:P] ~ dmnorm( mu.beta[], prior.T[,] )
    # gamma prior for the precision
    tau    ~ dgamma( 0.01, 0.01 )
    # deterministic calculation of variance
    s2    <- 1/tau
    s <-sqrt(s2)
}
INITS
list( tau=1, beta=c(1, 0, 0) )
DATA (RECT.)
l(n=25, P=3)
time[]  X[,2] X[,3]
 16.68   7 560
 ...   ... ...
 10.75   4 150
END
```

Results

- The model was run for $c^2 = n = 25$ and for $c^2 = 10^4$
- Results are presented in Tables 5.7 and 5.8, respectively.
- Results (especially those concerning the variance parameter σ^2) are sensitive to the choice of c^2, indicating that the unit information choice $c^2 = n = 25$ here is informative.

Table 5.7: WinBUGS posterior summaries for Example 5.1 after 2000 iterations and additional discarded 1000 burnin iterations using Zellner's g-prior ($c^2 = n = 25$)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	2.228	2.039	0.04583	-1.974	2.275	6.238	1001	2000
beta[2]	1.565	0.3246	0.007071	0.9216	1.563	2.205	1001	2000
beta[3]	0.01358	0.006783	1.423E-4	4.297E-5	0.01348	0.02702	1001	2000
s	6.247	0.938	0.01813	4.785	6.134	8.393	1001	2000

Table 5.8: WinBUGS posterior summaries for Example 5.1 after 2000 iterations and additional discarded 1000 burnin iterations using Zellner's g-prior ($c^2 = 10^4$)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	2.329	1.07	0.02317	0.1783	2.341	4.406	1001	2000
beta[2]	1.622	0.1695	0.004654	1.297	1.618	1.945	1001	2000
beta[3]	0.0143	0.003547	7.901E-5	0.007414	0.01427	0.0214	1001	2000
s	3.153	0.4811	0.01238	2.375	3.109	4.253	1001	2000

5.4 Analysis of variance models

- The normal models discussed in Sections 5.2 and 5.3 assess the association between continuous variables.
- Analysis of variance (ANOVA) models also assume a normal response variable, but now the explanatory variables are categorical.
- Here we will discuss one-way ANOVA models (with one categorical covariate).
- Also we will refer to their parametrization.

5.4.1 The one-way ANOVA model

- Assume a categorical variable A (also called *factor*) with levels $\ell = 1, 2, \ldots, L_A$ and a continuous response variable Y.
- When the categorical variable A is influencing the mean of $Y \Rightarrow$ different means of Y for each category of A.
- Thus the model can be summarized by

$$Y \sim N(\mu'_\ell, \sigma^2),$$

  - $\ell = 1, 2, \ldots, L_A$ indicates the group (category) of factor A
  - μ'_ℓ indicates the mean of Y for the ℓ category

Parametrization of the Mean Values

We can rewrite

$$\mu'_\ell = \mu_0 + \alpha_\ell \tag{6}$$

This expression decomposes the original mean of each category level μ'_ℓ to

- an overall common parameter μ_0 called *constant* and
- *group-specific* parameters α_ℓ, which are termed *effects* of ℓ level on the response variable Y

The interpretation of these parameters depends on the parametrization used for α_ℓ; see next subsection for details.

Full Model Specification

- Consider a random sample of n individuals resulting in n_ℓ subjects for each level ℓ ($\ell = 1, 2, \ldots, L_A$) of variable A.

  Then the model can be written as

$$Y_{\ell k} \sim N(\mu'_\ell, \sigma^2) \quad \text{and} \quad \mu'_\ell = \mu_0 + \alpha_\ell \tag{7}$$

  for $k = 1, 2, \ldots, n_\ell$ and $\ell = 1, 2, \ldots, L_A$.

- In practice, we usually observe n pairs (a_i, y_i) that are realizations of the random variables (A_i, Y_i), where $a_i \in \{1, 2, \ldots, L_A\}$ is the group or level at which the ith subject belongs.

  In this case the model can be rewritten as

$$Y_i \sim N(\mu_i, \sigma^2) \quad \text{and} \quad \mu_i = \mu'_{a_i} = \mu_0 + \alpha_{a_i} \tag{8}$$

  for $i = 1, 2, \ldots, n$.

5.4.2 Parametrization and parameter interpretation

- From (7) it is evident that we are interested in estimating the mean values μ'_ℓ of Y for each level of A.
- Thus, the original formulation $\mu_i = \mu'_{a_i}$ can be used to directly estimate the parameters of interest.
- Nevertheless, parametrization (8) is used for two reasons:
  1. It separates the constant overall effect from the effect of the categorical variable A,
  2. It allows for generalization of the ANOVA formulation when additional categorical explanatory variables are involved in the model.

- In the direct estimation of the mean values μ'_ℓ, we estimate L_A parameters (one for each group/level).
- When the alternative parametrization (7) is used, then we need to estimate $L_A + 1$ parameters.
- To make the model identifiable (i.e., the estimation feasible) and the two models equivalent, we impose one constraint on the new set of parameters.
- This constraint also specifies the interpretation and practical meaning of each parameter.
- Many parametrizations can be imposed by using different constraints, but two of them are most frequently met in statistical literature: the corner (CR) and the sum-to-zero (STZ) constraints.
- These two parametrizations are described here in detail.

5.4.2.1 Corner constraints.

- The effect of a level $r \in \{1, 2, \ldots, L_A\}$ is set equal to zero: $\alpha_r = 0$.
- This level r is referred to as the *baseline* or *reference category* of factor A.
- Usually the first or the last (in order) level is used as the reference category.
- In medicine, placebo or standard (old) treatment are used as baseline levels.
- In the following, we use the first level as the reference category: $\alpha_1 = 0$. Under this parametrization the mean of Y will be summarized by

$$\begin{aligned} \mu'_1 = E(Y|A=1) &= \mu_0 \\ \mu'_\ell = E(Y|A=\ell) &= \mu_0 + \alpha_\ell \quad \text{for} \quad \ell \geq 2, \end{aligned}$$

  - The constant parameter has a straightforward interpretation.
  - It is the mean of Y for the reference category.
  - Effect α_ℓ is the difference $\mu'_\ell - \mu'_1 = \alpha_\ell$.
  - Hence, α_ℓ is the expected difference of Y for an individual belonging in ℓ group in comparison to an individual from the reference group.

5.4.2.2 Sum-to-zero constraints.

- Under STZ, the following constraint is imposed:

$$\sum_{\ell=1}^{L_A} \alpha_\ell = 0. \tag{9}$$

- In practice, in the likelihood we substitute one parameter (usually the first or the last one) with the function resulting from the STZ constraint (9) e.g.

$$\alpha_1 = -\sum_{\ell=2}^{L_A} \alpha_\ell \ . \tag{10}$$

STZ Parameter Interpretation

- The interpretation is different from the corresponding interpretation of the CR parameters.
- In STZ, the constant term encapsulates an overall mean effect since

$$\sum_{\ell=1}^{L_A} \mu'_\ell \quad = \quad L_A\mu_0 + \sum_{\ell=1}^{L_A} \alpha_\ell = L_A\mu_0 \Leftrightarrow \mu_0 = \frac{1}{L_A}\sum_{\ell=1}^{L_A} \mu'_\ell,$$

- α_ℓ describes deviations of each level from this overall mean effect.
  - Positive values induce an increased effect in comparison to the overall mean
  - Negative values induce effects lower than the overall mean level

5.4.3 One-way ANOVA model in WinBUGS

- Here we assume the data given in pairs (a_i, y_i) referring to the characteristics of the ith individual.
- The stochastic part of the likelihood is the same as the one in normal regression models.
- The deterministic part of the likelihood is slightly changed since the mean must be specified as a function of each level of A [see Eq. (8)].

WinBUGS Code for ANOVA

- Hence, the likelihood is defined in WinBUGS

```
for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- mu0 + alpha[ a[i] ]
}
```

- The imposed constraint must be set outside the likelihood loop.
  - For CR parametrization, we set

```
a[1] <- 0.0
```

  - while, for STZ parametrization, we set

```
a[1] <- -sum( a[2:LA] )
```


WinBUGS Code for ANOVA - prior

- As a prior for μ and α_ℓ (for $\ell = 2, \ldots, L_A$), we consider a simple normal distribution with mean zero and low precision to express prior ignorance.
- Hence, in the WinBUGS code

```
mu0   ~ dnorm( 0.0 , 1.0E-4)
for (l in 2:LA){ alpha[i] ~ dnorm( 0.0, 1.0E-04 )}
```

  the prior for the precision τ is defined as in normal regression models.
- No prior is imposed on the constrained parameter α_1 since it is set equal to zero and therefore it does not appear in the likelihood equation.
- In CR parameterization, α_1 is a constant node
- In STZ parametrization, it is a logical/deterministic node since it is defined as a function of the remaining parameters.

Priors Under Different Parametrizations

- When using the prior setup described above, we must be very careful since, under different parametrizations, we impose different prior distributions on the group means μ'_ℓ.
- When the prior precision is small, differences due to this incompatibility of prior specification will be minor
- When prior information is used, these parameters must be specified carefully in order to lead to compatible prior beliefs.

5.4.4 A one-way ANOVA example using WinBUGS

Example 5.2. Evaluation of candidate school tutors.

- *The director of a private school wishes to employ a new mathematics tutor.*
- *The ability of four candidates is examined using a small study.*
- *A group of 25 students was randomly divided into four classes.*
- *In all classes, the same mathematical topic was taught for 2 hours per day for 1 week.*
- *After completing the short course, all students had to take the same test.*
- *Their grades were recorded and compared (see Table 5.9). The administrator wishes to employ the tutor whose students attained higher performance at the given test.*

Table 5.9: Data for Example 5.2 (school tutors' evaluation data)

Candidate	Students' grades
1	84 58 100 51 28 89
2	97 50 76 83 45 42 83
3	64 47 83 81 83 34 61
4	77 69 94 80 55 79

Setting up the data and the model code

- Data are coded in WinBUGS using two columns/variables
  1. Student's grades
  2. Which tutor was teaching
- Within the data we have also defined
  1. The number of cases ($n = 25$) and
  2. The number of tutors (`TUTORS`$= 4$).
- Table 5.10 provides the list format of the data.

Table 5.10: WinBUGS list format data for Example 5.2

```
list( n=25, TUTORS=4,
      grade=c(84, 58, 100, 51, 28, 89, 97, 50,  76, 83, 45, 42, 83,
              64, 47,  83, 81, 83, 34, 61, 77, 69,  94, 80, 55, 79),
      class=c(1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2,
              3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4) )
```

Model code

- The model can be defined according to the previous guidelines.
- Each group (tutor) mean = (constant term) + (tutor effect with index given by the variable `classes`).
- Initial values: for `mu` and `alpha` except from `alpha[1]`.
- The code and the initial values are provided in Table 5.11.
- STZ parametrization is used, (CR parametrization is provided as comments).

Table 5.11: WinBUGS code and initial values for Example 5.2 (For CR parametrization, remove `#` in line 10 and adding it to line 8 of the code).

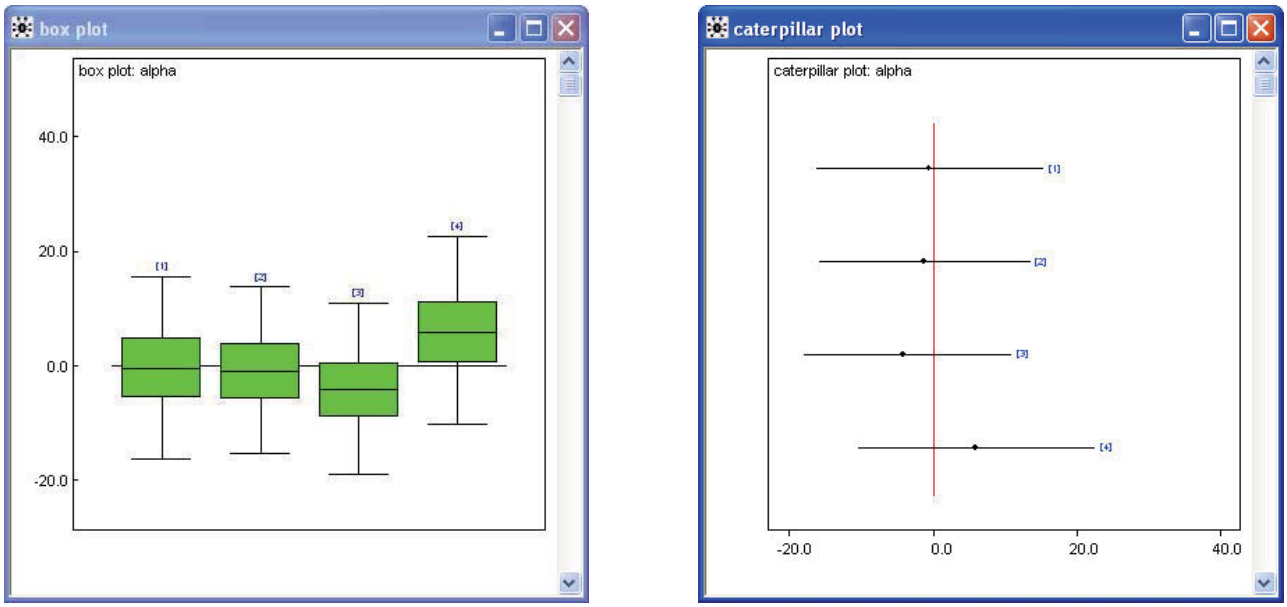
```
model{
    # model's likelihood
    for (i in 1:n){
        mu[i] <- m + alpha[ class[i] ]
        grade[i] ~ dnorm( mu[i], tau )
    }
    #### stz constraints
    alpha[1] <-  -sum(alpha[2:TUTORS])
    #### CR Constraints
    # alpha[1] <- 0.0
    # priors
    m~dnorm( 0.0, 1.0E-04)
    for (i in 2:TUTORS){ alpha[i]~dnorm(0.0, 1.0E-04)}
    tau ~dgamma( 0.01, 0.01)
    s <- sqrt(1/tau) # precision
}
INITS
list( m=1.0, alpha=c(NA, 0,0,0), tau=1.0 )
```

Results

- 1000 burnin and additional 2000 iterations
- Posterior summaries are given in Table 5.12
- We are interested in evaluating the overall performance of each tutor that is encapsulated by each parameter α_j (`alpha[j]`) for $j = 1, 2, 3, 4$.
- Fourth tutor has a higher performance (close to 6).
- The remaining tutors have negative effects, indicating that their performance is below the overall mean.
- Nevertheless, from Figure 5.2, the posterior distributions of α_j are not clearly discriminated, indicating that the between-tutor differences are minor.
- Since the school needs to hire only one tutor, we recommend hiring the last one but keeping in mind that differences in the tutors' performance in this small study did not indicate clear differences between tutors' actual abilities.

Table 5.12: Posterior summaries for ANOVA parameters of Example 5.2

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha[1]	-0.5661	8.016	0.1488	-16.27	-0.661	15.43	1001	2000
alpha[2]	-1.218	7.437	0.1442	-15.77	-1.317	13.6	1001	2000
alpha[3]	-4.17	7.323	0.1497	-18.04	-4.158	11.04	1001	2000
alpha[4]	5.955	8.345	0.1626	-10.34	6.066	22.56	1001	2000
m	68.96	4.561	0.1109	60.0	68.98	78.11	1001	2000
s	22.21	3.638	0.08999	16.54	21.74	30.55	1001	2000



(a) Boxplots (b) Error bars

Figure 5.2: Posterior boxplots and error bars for tutors' effects in Example 5.2 (school tutors' evaluation).

Two Way ANOVA

An example can be found in my book (Chapter 5, pages 173–184) with figures and interpretation for the model with and without interactions.

References

Fernandez, C., Ley, E. and Steel, M. (2000), "Benchmark priors for Bayesian model averaging", *Journal of Econometrics* **100**, 381–427.

Fouskakis, D., Ntzoufras, I. and Draper, D. (2009), "Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care", *Annals of Applied Statistics* **3**, 663–690.

Kass, R. and Wasserman, L. (1995), "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion", *Journal of the American Statistical Association* **90**, 928–934.

Liang, F., Paulo, R., Molina, G., Clyde, M. and Berger, J. (2008), 'Mixtures of g priors for Bayesian variable selection', *Journal of the American Statistical Association* **103**, 410–423.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Vol. 37, 2nd ed., Chapman & Hall, Cambridge, UK.

Montgomery, D. and Peck, E. (1992), *Introduction to Regression Analysis*, Wiley, New York.

Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, Wiley Series in Computational Statistics, Hoboken, NJ.

Ntzoufras, I. (2010), "Bayesian analysis of the normal regression model", in K. Bocker, ed., *Rethinking Risk Measurement and Reporting: Uncertainty, Bayesian Analysis and Expert Judgement*, Risk Books (In Press). ISBN-10: 1-906348-40-5, ISBN-13: 978-1-906348-40-3.

Raftery, A., Madigan, D. and Hoeting, J. (1997), "Bayesian model averaging for linear regression models", *Journal of the American Statistical Association* **92**, 179–191.

Ryan, T. (1997), *Modern Regression Methods*, Wiley Series in Probability and Statistics, Wiley, New York.

Spiegelhalter, D., Abrams, K. and Myles, J. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Statistics in Practice, Wiley,

Chichester, UK.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002), "Bayesian measures of model complexity and fit (with discussion)", *Journal of the Royal Statistical Society B* **64**, 583–639.

Stanton, J. (2001), "Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors", *Journal of Statistics Education* **9**(3).

Zellner, A. (1986), "On assessing prior distributions and Bayesian regression analysis using g-prior distributions", in P. Goel and A. Zellner, eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, North-Holland, Amsterdam, pp. 233–243.