BAYESIAN MODELING USING WINBUGS
An Introduction

Ioannis Ntzoufras
Department of Statistics,
Athens University of Economics and Business
CHAPTER 5

NORMAL REGRESSION MODELS

5.1 GENERAL MODELING PRINCIPLES.

Statistical models are nowadays used to describe parsimoniously real life problems observed under uncertainty. A statistical model is a collection of probabilistic statements (and equations) which describe and interpret present or predict future performance. It consists of three important components:

1. the response variable (or variables) $Y$,
2. the explanatory variables $X_1, X_2, \ldots, X_p$ and
3. a linking mechanism between the two set of variables.

The response variables $Y$ are the main study variables and they compose the stochastic part of the model. By the term stochastic we refer to random variables whose outcome is uncertain before it is observed. Concerning these variables, we are frequently interested to describe the mechanism underlying or leading to the appearance of a certain outcome of $Y$ and predict a future outcome of $Y$. Since the response variable is the stochastic component of the model, we can write that

$$ Y \sim \text{Distribution}(\theta) $$

where $\theta$ is the parameter vector of the distribution used. For example, for a normal regression model, the response-stochastic component of the model is written as

$$ Y \sim \text{Normal}(\mu, \sigma^2). $$
NORMAL REGRESSION MODELS

Models with one response variable are called univariate while models with more than one response variables are called multivariate. In this book we will focus attention in univariate models.

As explanatory variables $X_1, \ldots, X_p$, we consider all variables which possibly influence the response variable $Y$. Inference concerning the significance, the type (negative or positive) and the magnitude of the effect of each $X_i$ on $Y$ is the main focus in such models. Usually $X_i$ are considered as fixed, non stochastic components i.e. deterministic nodes in WinBUGS. Hence, it is more precise to define the distribution of $Y$ conditional on the observed explanatory variables

$$Y | X_1, \ldots, X_p \sim \text{Distribution}\left(\theta(\beta, \phi, X_1, \ldots, X_p)\right).$$

Parameter vector $\theta$ is expressed as a function of the explanatory variables and a new alternative set of parameters $(\beta, \phi)$ which substitutes the original ones in terms of estimation and inference. Concerning the new set of parameters, vector $\beta$ summarizes the association between the response and the explanatory variables while $\phi$ 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 be also estimated via the explanatory variables. The function used to connect the stochastic and the deterministic part of the model (variables $Y$ and $X_i$’s) can be called as the ‘generalized linking’ function. The above terminology and principles were originally introduced for the definition of generalized linear models McCullagh & Nelder (1989) but they can be adopted for a wide range of models. For example, the simpler case of the above general linking function is to express the mean of the response variable as a function of the linear combination of the explanatory variables, hence we can write

$$\mu = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} \beta_j X_j\right).$$

In the above equation, the linear combination of all explanatory variables is used to predict the expected value of the response variable $Y$ and is often called the linear predictor $\eta$ of the model. The above set-up is introduced in generalized linear models and, within this context, $g(\mu)$ is referred as the ‘link’ function.

To complete the above formulation, a prior distribution must be defined for the parameters under estimation. Thus, in the above formulation a prior distribution $f(\beta, \phi)$ for the parameters $(\beta, \phi)$ remains to be specified.

Finally, explanatory variables are usually defined as deterministic (fixed) quantities. In practice these variables are frequently random. Within the general framework of statistics it is generally easy to extend the model by either simply considering $X_i$’s as random variables with additional parameters under estimation; see Ryan (1997, p.34-35) for a discussion. This can be specified in a straightforward manner within the Bayesian framework using additional hierarchical levels in our model (see chapter 7 for details).

5.2 MODEL SPECIFICATION IN NORMAL REGRESSION MODELS

Normal regression models are the most popular models in statistical science. They are based on the initial work of Sir Francis Galton in the late years of the 19th century (Stanton 2001). In regression models, the response variable $Y$ is considered to be a continuous
random variable defined in the whole set of real numbers following the normal distribution 
with mean \( \mu \) and variance \( \sigma^2 \). Therefore the model can be summarized by the following 
equations

\[
Y \sim \text{Normal}(\mu(\beta, X_1, \ldots, X_p), \sigma^2)
\]

with

\[
\mu(\beta, X_1, \ldots, X_p) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p = \beta_0 + \sum_{j=1}^{p} \beta_j X_j
\]

where \( \sigma^2 \) and \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \) is the set of regression parameters under estimation. 
Frequently, the following alternative representation of the regression model is adopted

\[
Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon; \quad \varepsilon \sim \text{Normal}(0, \sigma^2).
\]

Although the above formulation has a nice interpretation since the response variable is 
directly expressed as a function of the explanatory variables plus a random normal error 
with variance \( \sigma^2 \), the initial model expression given by (5.1) and (5.2) is more general and 
follows the model building principles described in section 5.1.

5.2.1 Specifying the likelihood

Let us observe a sample of size \( n \) with response values \( y = (y_1, \ldots, y_n)^T \) and \( x_{i1}, \ldots, x_{ip} \) 
the values of the explanatory variables \( X_1, \ldots, X_p \) for individuals \( i = 1, \ldots, n \). Then the 
model is expressed as

\[
Y_i \sim \text{Normal}(\mu_i, \sigma^2)
\]

\[
\mu_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \quad \text{for } i = 1, \ldots, n.
\]

Within WinBUGS the normal distribution is defined in terms of its precision \( \tau = \sigma^{-2} \). 
Hence the above likelihood in WinBUGS will be written as commands

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

The last two commands are used to deterministically specify the connection between the 
variance, the standard deviation and the precision parameter \( \tau = \sigma^{-2} \) used by the normal 
distribution in WinBUGS. Moreover, in the above code all parameters \( \beta_j \) are defined 
separately as single scalar nodes and while explanatory variables are specified as vector 
nodes with names \( x_1, \ldots, x_p \) of length \( n \). When monitoring these parameters, each one 
of them must be set separately in the sample monitor tool of WinBUGS.
5.2.2 Specifying a simple independent prior distribution.

In Normal models, the simplest approach is to assume that all parameters are a-priori independent having the following structure.

$$f(\beta, \tau) = \prod_{j=0}^{p} f(\beta_j) f(\tau).$$

with

$$\beta_j \sim \text{Normal}(\beta_{0j}, c_j^2) \text{ for } j = 0, \ldots, p \text{ and } (5.4)$$

$$\tau \sim \text{Gamma}(a, b).$$

The above gamma prior of the precision parameter induces prior mean and variance given by

$$E(\tau) = \frac{a}{b} \text{ and } Var(\tau) = \frac{a}{b^2},$$

respectively. In the above prior set-up, we have substituted the variance $\sigma^2$ by the corresponding precision parameter $\tau$ in order to make it compatible to the WinBUGS notation. The above gamma prior corresponds to an inverse gamma prior distribution for the original variance parameter with prior mean and variance given by

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

respectively.

When no information is available, a usual choice for the prior mean is the zero value ($\beta_{0j} = 0$). This prior choice centers our prior beliefs around zero which corresponds to the assumption of no effect of $X_j$ on $Y$. By this way, we express our prior doubts about the effect of $X_j$ on $Y$ and, for this reason, such prior is called ‘sceptical’ prior by Spiegelhalter, Abrams & Myles (2004, p.90 & p.158–160). The prior variance $c_j^2$ of the effect $\beta_j$ is set equal to a large value (for example $10^{-4}$) to represent high uncertainty or prior ignorance. Similarly, for $\tau$ we use equal low prior parameter values setting by this way its prior mean equal to one and its prior variance large. For example, we may use $a = b = 0.01$ which results in $E(\tau) = 1$ and $V(\tau) = 100$. This approach is also adopted in all illustrations of the WinBUGS manual and example volumes. More details concerning the specification of more complicated prior distributions are provided in sections 5.3.2 and 5.3.3 which follow.

Within WinBUGS, the above prior setup can be incorporated by simply adding

```plaintext
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 )
```

In the above syntax, value $1.0E-4$ is the scientific notation for $1.0 \times 10^{-4} = 0.001$ which is the prior precision of each $\beta_j$ and corresponds to prior variance equal to $10^4$. The above definition is 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 regression coefficient refers to the effect of explanatory variable $X_j$ on the expectation of the response variable $Y$ adjusted for the rest of covariates. The inference concerning the model parameter can be divided in three basic stages

- Is the effect of $X_j$ important for the prediction or description of $Y$?
- What is the association between $Y$ and $X_j$ (positive, negative or other)?
- What is the magnitude of the effect of $X_j$ on $Y$?

Concerning the first query, we initially focus on examining whether the posterior distribution of $\beta_j$ is scattered around zero (or not). Posterior distributions far away from the zero value will indicate an important contribution of $X_j$ on the prediction of the response variable. Although, formal Bayesian hypothesis testing is not based on simply examining the posterior distribution and their credible intervals, such analysis can offer a first and reliable tool to trace important variables.

In the second stage, we identify whether the relationship is positive or negative. This can be based on the signs of the posterior summaries of central and relative location (for example 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 means that changes of the explanatory variable $X_j$ cause changes of the same direction for variable $Y$ while negative association means that changes of the explanatory variable $X_j$ cause changes of the opposite direction for variable $Y$. Within this analysis, we can a-posteriori calculate the posterior probability

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

When the zero value lies at the center of the posterior distribution then the above value will be close to $1/2$ indicating that there is no clear positive or negative effect of $X_j$ on $Y$. When $\pi_0$ is low (for example lower than 2.5%, 1% or 0.5%) then we may conclude positive or negative association depending on the sign of the posterior location summaries. Within WinBUGS we can calculate the posterior probability $f(\beta_j > 0|y)$ using the syntax

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

which creates a binary node $p.betaj$ taking values equal to one when $\beta_j$ is positive and zero otherwise. Obtaining the posterior mean via the sample monitor tool provides us the estimate of the posterior probability $f(\beta_j > 0|y)$.

In WinBUGS, it is also convenient to calculate the deviance information criterion (DIC, Spiegelhalter et al. 2002) to compare models with different covariates and, by this way, evaluate their importance concerning their effect on $Y$. Brief description of DIC as well as illustration of its calculation in WinBUGS is provided in section 2.9.1.5 of chapter 2. In order to use DIC in regression models, we need to fit models including and excluding the variable of interest and then select the one with the lower value of DIC. In the case of a large number of covariates, this procedure can be quite tedious since a large number of model must be fitted before concluding to the model with the lowest value of DIC. More formal approaches concerning model checking, comparison and selection are described in chapters 8 and 9.

Finally, the magnitude of the effect of variable $X_j$ on $Y$ is given by the posterior distribution itself. To simplify interpretation, we may focus on the posterior mean (or alternatively
the median) as a point estimate for the magnitude of this effect. Hence, an increase of one unit of $X_j$, given that the rest of the covariates will remain stable, will induce an a-posteriori average change on the expectation of $Y$ equal to posterior mean of $\beta_j$.

Concerning the constant parameter $\beta_0$, its interpretation corresponds to the expected value of the response variable $Y$ when the observed values of all covariates are equal to zero. Frequently such combination lies outside the range of the observed covariate values. In such cases, the interpretation of $\beta_0$ is not reliable since we infer or predict the behavior of $Y$ for values of $X_j$’s that have not been observed. Frequently, direct interpretation of $\beta_0$ does not lead to realistic and sensible interpretation. An alternative is to center around zero all explanatory variables $X_j$ by subtracting their sample mean. In this case, the constant $\beta_0^c$ represents the expected value of $Y$ when all covariates are equal to its sample means representing by this way the expected response $Y$ for an “average” or “typical” subject according to our sample. In WinBUGS this quantity can be directly estimated using the command

$$\text{typical.y} \leftarrow \text{beta0} + \text{beta1} \times \text{mean(x1[])} + \ldots + \text{betap} \times \text{mean(xp[])}$$

without changing the parameterization of the original model. The above approach can be also used for the calculation of the expected values of $Y$ for any combination of values of $X_j$’s.

Parameter precision $\tau$ (and the variance $\sigma^2$) indicates how precise is the model. If the precision $\tau$ is high ($\sigma^2$ low) then the model can predict (or describe) with precision the expected values of $Y$. Therefore, we can rescale this quantity using the sample variance of the response variable $Y$, $s_Y^2$, by using the $R^2_B$ statistic given by

$$R^2_B = 1 - \frac{\tau^{-1}}{s_Y^2} = 1 - \frac{\sigma^2}{s_Y^2},$$

where $s_Y^2$ is the sample variance of $Y$. The above quantity can be interpreted as the proportional reduction of uncertainty concerning the response variable $Y$ achieved by incorporating the explanatory variables $X_j$ in the model. Moreover, it can be thought as the Bayesian analogue of the adjusted coefficient of determination $R^2_{adj}$ (used in frequentistic approach of normal regression model) given by

$$R^2_{adj} = 1 - \frac{\hat{\sigma}^2}{s_Y^2},$$

where

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \text{ with } \hat{y}_i = \hat{\beta}_0 + \sum_{i=1}^{p} X_{ij}\hat{\beta}_j$$

where $\hat{\beta}_j$ are the maximum likelihood estimates of $\beta_j$.

In order to calculate $R^2_B$ in WinBUGS we can use the commands

```
sy2 <- pow(ad(y[]), 2)
R2B <- 1 - a2/sy2
```

or directly using the precision parameter $\tau$ and the syntax

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

### 5.2.4 A regression example using WinBUGS
EXAMPLE 5.1

**Soft Drink Delivery Times.** The following example deals with the quality of the delivery system network of a soft drink company; see example 4.1 in Montgomery & Peck (1992). In this problem, interest lies in the estimation of the required time needed by each employee to refill an automatic vending machines owned and served by the company. For this reason, a small quality assurance study was set-up by an industrial engineer of the company. As the response variable he considered the total service time (measured in minutes) of each machine including its stocking with beverage products and any required maintenance or housekeeping. After examining the problem, the industrial engineer recommended two important variables which affect delivery time:

1. the number of cases of stocked products and
2. the distance walked by the employee (measured in feet).

A dataset of 25 observations was finally collected.

**Setting up the data and the model code.** Following the approach described in sections 5.2.1 and 5.2.2, we define the data either in a rectangular or in a list format. The rectangular format of the data is provided in Table 5.1., while the full model code of the example including the list data format and the initial values is given in Table 5.2.. All three variables used in the model (time, cases, distance) are defined as separate vectors in the list data format while, in the initial values, each parameter $\tau$, $\beta_0$, $\beta_1$ and $\beta_2$ was initialized separately.

**Results.** Posterior summaries and densities, after running the MCMC algorithm for 3000 iterations and discarding the initial 1000 ones, are provided in Table 5.3. and Figure 5.1. respectively. 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.

Concerning the posterior distribution of $\sigma$, we observe that with the current model we can predict the expected delivery time with with an a-posteriori expectation equal to 3.4 minutes.

Considering as point estimates the posterior means, we end up with model

$$\text{Expected Time} = 2.36 + 1.6 \times \text{Cases} + 0.015 \times \text{Distance}.$$  

Minor changes are observed in the above equation if posterior medians are used as point estimates instead.

Observing all parameters, we can infer that 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.

Furthermore, for each additional case stocked by the employee, the expected delivery time is a-posteriori expected to increase by 1.6 minutes (96 seconds). The increase of the expected delivery time for each additional case, lies between 1.3 and 2.0 minutes (76 and 118 seconds) with probability 95%. For every increase of the walking distance by one foot, the delivery time is a-posteriori expected to increase by 0.87 seconds while every 100 feet
<table>
<thead>
<tr>
<th>time[]</th>
<th>cases[]</th>
<th>distance[]</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.68</td>
<td>7</td>
<td>560</td>
</tr>
<tr>
<td>11.5</td>
<td>3</td>
<td>220</td>
</tr>
<tr>
<td>12.03</td>
<td>3</td>
<td>340</td>
</tr>
<tr>
<td>14.88</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>13.75</td>
<td>6</td>
<td>150</td>
</tr>
<tr>
<td>18.11</td>
<td>7</td>
<td>330</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>17.83</td>
<td>7</td>
<td>210</td>
</tr>
<tr>
<td>79.24</td>
<td>30</td>
<td>1460</td>
</tr>
<tr>
<td>21.5</td>
<td>5</td>
<td>605</td>
</tr>
<tr>
<td>40.33</td>
<td>16</td>
<td>688</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>215</td>
</tr>
<tr>
<td>13.5</td>
<td>4</td>
<td>255</td>
</tr>
<tr>
<td>19.75</td>
<td>6</td>
<td>462</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>448</td>
</tr>
<tr>
<td>29</td>
<td>10</td>
<td>776</td>
</tr>
<tr>
<td>15.35</td>
<td>6</td>
<td>200</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>132</td>
</tr>
<tr>
<td>9.5</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>35.1</td>
<td>17</td>
<td>770</td>
</tr>
<tr>
<td>17.9</td>
<td>10</td>
<td>140</td>
</tr>
<tr>
<td>52.32</td>
<td>26</td>
<td>810</td>
</tr>
<tr>
<td>18.75</td>
<td>9</td>
<td>450</td>
</tr>
<tr>
<td>19.83</td>
<td>8</td>
<td>635</td>
</tr>
<tr>
<td>10.75</td>
<td>4</td>
<td>150</td>
</tr>
</tbody>
</table>

**Table 5.1.** Rectangular WinBUGS format of “Soft Drink Delivery Times” example (example 5.1).
model{
  # model's likelihood
  for (i in 1:n){
    time[i] ~ dnorm( mu[i], tau ) # stochastic component
    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, 210, 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 ) )

Table 5.2. Full model code for “Soft Drink Delivery Times” example (example 5.1).
of additional walking distance increases by 1.5 minute the posterior expected delivery time (ranging between 0.7 and 2.2 minutes with probability 95%). In terms of meters, for every 100 meters of walking distance will increase the expected delivery time by 4.7 minutes on posterior average (one foot is equal to 0.3048 meters resulting in an increase of the expected delivery time by $100/0.3048 \times 0.01447 = 4.747$ minutes).

Parameter $\beta_0$ has no sensible interpretation in this example since the zero value is non-sense 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. We only observe that the zero value lies at the left tail of the posterior distribution within the range of the 95% posterior interval. Moreover, the posterior probability of positive $\beta_0$ is equal to 97.4%.

Since, the interpretation for $\beta_0$ is meaningless, focus can be given on the predicted value for the a typical or representative delivery route. According to the posterior summaries of node 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 MODELS.

#### 5.3.1 Defining the model using matrices.

The regression model described in the previous sub-section can be defined in WinBUGS using vectors and matrices instead of scalar nodes. To achieve that, within the likelihood loop, we substitute the mean specification code line with the following syntax

$$\mu[i] <- \beta_0 + \beta[1] \times x[i,1] + \ldots + \beta \times x[i,p]$$

Matrix $x$ (denoted by $x[,]$ in WinBUGS) is a $n \times p$ matrix. Each each column $x_j$ of $x$ corresponds to each explanatory variable $X_j$ while each row $x_{(i)}$ corresponds to the explanatory variable values of the $i$-th subject of the sample. Concerning the data definition, each column of the matrix can be defined using the rectangular data format with header $x[,1] \ldots x[,p]$. Otherwise, in list format, $x$ must be defined as an array with dimensions $n$ and $p$.

Moreover, when the number of variables $p$ is large, it is convenient to use the command $\text{inprod}(\ b[,] \ , \ x[,])$ to express the linear combination of the explanatory variables $\sum_{j=1}^{p} \beta_j x_{ij}$. Hence, the mean can be defined by the syntax

### Table 5.3.
WinBUGS posterior summaries for example 5.1 after 2000 iterations and additional discarded 1000 burn-in iterations.
Figure 5.1. Posterior densities of model parameters for example 5.1 (Soft Drink Delivery Times).

\[ \text{mu}[i] \leftarrow \text{beta0} + \text{inprod(b[], x[i,])} \]

where \( b = \beta_{\{0\}} = (\beta_1, \ldots, \beta_p) \).

Usually we incorporate the constant term in a matrix \( X = [1_n, x] \) of dimension \( n \times (p + 1) \) which called the data matrix; where \( 1_n \) is a vector of length \( n \) with all elements equal to one corresponding to the ‘constant’ term. The linear predictor in (5.3) can be now written as

\[ \mu_i = \beta X_{(i)} \]

where \( X_{(i)} \) is the \( i \)-th row of \( X \). This expression can be coded in WinBUGS by

\[
x[i,1]<-1
\]

\[
\text{mu}[i] \leftarrow \text{inprod(b[], X[i,])}
\]

In the above, \( X[,1] \) is a matrix of dimension \( n \times (p + 1) \) with all components of the first column \( x[,1] \) equal to one. Using this definition, \( \text{beta[1]} \) represents the constant coefficient \( \beta_0 \), \( X[,j] \) refers to the vector of values of \( X_{j-1} \) and \( \text{beta[j]} \) corresponds to coefficient \( \beta_{j-1} \) for \( j = 1, \ldots, p+1 \). In the rectangular format, data are defined in the usual manner with header names referring to columns from 2 to \( p+1 \), i.e. \( X[,2] \ldots X[,p+1] \). This approach enables us to monitor all regression coefficients simultaneously by simply considering the vector node \( \text{beta} \).

5.3.2 Prior distributions for Normal regression models

Conjugate analysis for normal regression model has been presented in section 1.5.5. The conjugate prior for the normal regression model is considered if we specify \( [\beta, \sigma^2] \) to a-priori follow a Normal-inverse-gamma distribution. Hence we can write

\[
\beta | \sigma^2 \sim N_P(\beta_0, c^2V\sigma^2) \quad \text{and} \quad \sigma^2 \sim IG(a, b)
\]  

(5.5)
where \( c^2 \) is a parameter controlling the overall magnitude of the prior variance.

Special case of the above family of prior distributions is popular Zellner’s (1986) \( g \)-prior in which

\[
V = (X^T X)^{-1}.
\]

Parameter \( c^2 \) was denoted by \( g \) in Zellner’s original publication. The default choice of \( c^2 = n \) is usually adopted since it has an interpretation of adding prior information equivalent to one data point (see for details in Kass & Wasserman 1995, Fouskakis, Ntzoufras & Draper 2006). This prior has been widely used because it considerably simplifies posterior computations and reduces the number of prior variance parameters that remain to be specified down to one.

Generally, the above conjugate prior set-up is very convenient for implementing Bayesian variable selection; see Raftery, Madigan & Hoeting (1997). Moreover, Zellner’s \( g \)-priors were widely used within this context since they allow us for a sensible default choice of prior distributions; see Fernandez, Ley & Steel (2000) for comparison between different values of \( c^2 \) and Liang, Paulo, Molina, Clyde & Berger (2006) for discussion and extensions concerning the \( g \)-priors.

When no information is available, we may simplify the above prior set up by setting the matrix

\[
V = c^2 I_P
\]

with \( P = p + 1 \) and \( c^2 \) set large to express prior ignorance (for example \( c = 100 \)). This induces that the components of the vector \( \beta \) will be a-priori independent. Hence we can simply rewrite the prior as

\[
\beta_j | \sigma^2 \sim N_P(\beta_{0j}, c^2 \sigma^2) \quad \text{for} \quad j = 0, 1, \ldots, p
\]  

(5.6)

where \( \beta_{0j} \) are the components of the prior mean vector \( \beta_0 \).

Another alternative is to consider the simpler prior set-up of section 5.2.2 where all parameters are a-priori independent. This choice is usually selected when no information is available. It is not conjugate, and hence MCMC methods need to be implemented in order to estimate the posterior distribution. Nevertheless, this prior set up is conditionally conjugate resulting in conditional posterior distributions for \( \beta \) and \( \tau \) that can be calculated analytically allowing us to construct an efficient Gibbs sampler.

Finally, alternative prior distributions on \( \beta \) have been proposed in the literature. For example a student’s t-distribution or Cauchy distribution can be used instead but usually no obvious differences are observed in the case that no prior information is available.

### 5.3.3 Multivariate normal priors in WinBUGS

In the case that vector nodes are used in the model specification, the independence prior (5.4) can be defined in WinBUGS by specifying the priors for \( \beta_j \)’s within a loop. Hence the \( P = p + 1 \) lines which define the prior distribution for each component of \( \beta \) can be substituted by

\[
\text{for } (j \in 1:P) \{ \beta[j] \sim \text{dnorm}(0.0, 1.0 \times 10^{-4}) \}
\]

If we wish to use the prior distribution (5.6) which a conjugate prior assuming that the elements of \( \beta \) vector are a-priori independent then we use the following the syntax

\[
\text{for } (j \in 1:P) \{ \beta[j] \sim \text{dnorm}(\beta_0[j], \tau / c^2) \}
\]
In order to define the conjugate prior (5.5), we firstly need to specify in WinBUGS the precision matrix used in the multivariate normal prior distribution. The elements of the prior precision matrix $T$ can be expressed as

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

(5.7)

for $l, j \in \{1, 2, \ldots, p+1\}$; where $V^{-1}$ is the inverse of matrix $V$ and $[V^{-1}]_{lj}$ is the $i$th row and $j$th column element of $V^{-1}$. Hence, in WinBUGS, we need to firstly calculate the $V^{-1}$ and then the elements of the prior precision matrix using (5.7) within a double for-loop. The multivariate prior can be specified by the following syntax:

```r
# calculation of the inverse matrix of V
# calculation of the elements of prior precision matrix
for(l in 1:P){ for (j in 1:P){
}}
# multivariate prior for the beta vector
beta[1:P] ~ dmnorm( beta0[], prior.T[,] )
# gamma prior for the precision
tau ~ dgamma( 0.01, 0.01 )
# deterministic calculation of variance
s2 <- 1/tau
```

In the above syntax $P$ stands for the length of $\beta$. The prior values $V$, $\beta_0$ and $c^2$ can be specified either within the list of the data or directly within the WinBUGS model code. For example the syntax

```r
c2 <- 100
for (j in 1:P){ beta0[j] <- 0.0 }
for (l in 1:P){ for (j in 1:P){
    V[l,j] <- equals(l,j)
}}
```

defines $\beta_0 = 0_P$, $c^2 = 100$ and $V = I_P$ resulting in a $\beta_j \sim N(0, 100\sigma^2)$ prior. In the above syntax, $V[l,j]$ will take the value one if $l = j$ (hence we have a diagonal element) and zero otherwise. Generally, it is easier to specify the directly the components of $T$ directly in the data than compute them inside the WinBUGS code.

Finally, Zellner’s g-prior can be specified using the above syntax but matrix $V^{-1}$ will be directly calculated using the syntax

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

The above index calculates matrix $A = V^{-1}$ with elements $A_{lj} = [V^{-1}]_{lj} = \sum_{i=1}^n x_i x_{ij}$ which are the elements of matrix $(X^T X)$ used in Zellner’s g-prior.

### 5.3.4 Example 5.1 (continued)

Here we rerun the same model using the Zellner’s g-prior. The basic code of the model is given in Table 5.4, in which matrix $V = (X^T X)$ is defined within the model code.
first column $X[,1]$ of matrix $X$ is defined within the model code while the rest of them within the data part. In Table 5.4., the first and the last rows of the data are also given (in a rectangular data format) while the rest of them were substituted 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). Alternatively, we can use the list format directly (with the help of R or Splus as described in section 2.4.6.2). The list format which can be used alternatively in code 5.4. is provided in Table 5.5.. As you may observe, some values (corresponding to the first column) are equal to $\text{NA}$ i.e. are defined as missing values. This is because the values of the first column are defined within the model’s code. Generally, if the list format is used, it is more convenient to specify the whole matrix $x$ within the data format and remove command $x[1,1] \leftarrow 1.0$ (line 19 of Table 5.4.) from the code.

In the case that we wish to define the prior parameters $V$, $\beta_0$ and $c^2$ in the data structure, the model code must be slightly changed by omitting the first lines appearing in the code of Table 5.4. related to the definition of the corresponding parameters. Code for the general normal-inverse gamma prior setup having all prior parameters specified in a data list format is provided in table 5.6.

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

### 5.4 Analysis of Variance Models

The above normal models assess the association between continuous variables. To be more specific, normal regression models identify which and how specific continuous explanatory variables influence a continuous response variable. Analysis of variance models, also assume a normal response variable but now the explanatory variables are categorical. In the following we provide specific examples with one and two variables as well as details concerning their parametrization. A short discussion for multifactor analysis of variance closes this section while an example of a three way analysis of variance model is provided in the next section.

#### 5.4.1 The one way ANOVA model

Let us assume a categorical variable $A$ (also called factor) with levels $\ell = 1, 2, \ldots, L_A$ and a continuous response variable $Y$. When we assume that the effect of the categorical variable $A$ is influencing the mean of the continuous variable $Y$, then this is equivalent to defining different means of $Y$ for each category of $A$. Thus, assuming normal distributions for the response variable $Y$, the model can be summarized by

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

where $\ell = 1, 2, \ldots, L_A$ is indicating the group (category) of factor $A$ from which $Y$ originates from and $\mu'_\ell$ indicates the mean of $Y$ for the $\ell$ category. An alternative way is to write

$$\mu'_\ell = \mu_0 + \alpha_\ell$$

The above expression decomposes the original mean of each category level $\mu'_\ell$ to an overall common parameter $\mu_0$ which is called constant and group specific parameters $\alpha_\ell$ which
model{

  # definition of prior parameters
  c2 <- 10000
  # prior means
  for (j in 1:P) { beta0[j] <- 0.0 }
  # calculation of xtx
  for (i in 1:P){ for (j in 1:P){
  }}
  # calculation of the elements of prior precision matrix
  for(l in 1:P){ for (j in 1:P){
  }}

  # 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 component
    # link and linear predictor
    mu[i] <- inprod( beta[], X[i,] )
  }

  # prior distributions
  # -------------------
  # calculation of the inverse matrix of V
  # prior parameters
  # multivariate prior for the beta vector
  beta[1:P] ~ dmnorm( beta0[], 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.)
list(n=25, P=3)
time[ ] X[,2] X[,3]
16.68 7 560
... ... ...
10.75 4 150
END

Table 5.4. WinBUGS code for Example 5.1 using Zellner's g-prior and parameter vectors (data are suppressed to save space).
NORMAL REGRESSION MODELS

are called effects of $\ell$ level on the response variable $Y$. The interpretation of the above parameters depends on the parametrization used for $\alpha_\ell$; see next sub-section for details.

Let us now consider a random sample of $n$ individuals resulting to $n_\ell$ subjects for each level $\ell$ ($\ell = 1, 2, \ldots, L_A$) of variable $A$. Then the above model can be written as

$$y_{\ell k} \sim N(\mu_\ell', \sigma^2) \text{ and } \mu_\ell' = \mu_0 + \alpha_\ell$$

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

In practice, we usually observe $n$ pairs $(a_i, y_i)$ which are realizations of the random variables $(A_i, Y_i)$; where $a_i \in \{1, 2, \ldots, L_A\}$ is the group/level in which the $i$-the subject belongs. In this case the above model can be written as

$$y_i \sim N(\mu_i, \sigma^2) \text{ and } \mu_i = \mu_{a_i}' = \mu_0 + \alpha_{a_i}$$

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

5.4.2 Parametrization and parameter interpretation

From (5.10) it is evident that we are interested to estimate the mean values $\mu_\ell'$ 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, parameterization (5.9) is used for two reasons: firstly it separates the constant overall effect from the effect of the categorical variable $A$ and secondly it allows for generalization of the above formulation when additional categorical explanatory variables are involved in the model.

In the direct estimation of the mean values $\mu_\ell'$, we estimate $L_A$ parameters (one for each group/level). When the alternative parametrization (5.9) 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 the practical meaning of each parameter. Many parameterizations 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 parameterizations are described here in detail.

Corner constraints. In corner constraints, the effect of a level $r \in \{1, 2, \ldots, L_A\}$ is set equal to zero:

$$\alpha_r = 0.$$
model{
    # definition of prior parameters
    # calculation of the inverse matrix of V
    # calculation of the elements of prior precision matrix
    for (l in 1:P){ for (j in 1:P){
    }}
    # model's likelihood
    # ------------------
    for (i in 1:n){
        # specifying the constant term in the first column
        time[i] ~ dnorm( mu[i], tau ) # stochastic component
        # link and linear predictor
        mu[i] <- inprod( beta[], X[i,] )
    }
    # prior distributions
    # # ------------------
    # calculation of the inverse matrix of V
    # prior parameters
    # multivariate prior for the beta vector
    beta[1:P] ~ dmnorm( beta0[], 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) )
list( n=25, P=3, c2=25, 
      beta0=c(0,0,0),
      V=structure(.Data = c( 0.113215186112351, -0.0044485932353407, 
                             -8.3672569807385E-05, -0.0044485932353407, 
                             0.00274378329085448, -4.78570865728707E-05, 
                             -8.36725698073853E-05, -4.78570865728707E-05, 
                             1.22874474243973E-06), .Dim = c(3,3)),
      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),
      X=structure(.Data=c(1, 7, 560, 1, 3, 220, 1, 3, 340, 1, 4, 
                          80, 1, 6, 150, 1, 7, 330, 1, 2, 110, 1, 7, 210, 1, 30, 1460,
                          1, 5, 605, 1, 16, 688, 1, 10, 215, 1, 4, 255, 1, 6, 462,
                          1, 9, 448, 1, 10, 776, 1, 6, 200, 1, 7, 132, 1, 3, 
                          36, 1,17, 770, 1, 10, 140, 1, 26, 810, 1, 9, 450, 1, 8, 
                          635, 1, 4, 150), .Dim = c(25,3)) )

Table 5.6. WinBUGS code for Example 5.1 using multivariate conjugate Normal-gamma prior (all prior parameters are specified within the list data format; data for matrix V are defined within the list data format).
This level \( r \) is referred as the baseline or reference category of factor \( A \). Usually the first or the last (in order) level is used as 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 i.e. \( \alpha_1 = 0 \). Under this parameterization the mean of \( Y \) will be summarized by

\[
\mu'_\ell = E(Y|A=1) = \mu_0 \\
\mu'_\ell = E(Y|A=\ell) = \mu_0 + \alpha_\ell \text{ for } \ell \geq 2.
\]

From the above, it is obvious that the constant parameter has a straightforward interpretation. It is simply the mean of \( Y \) for the reference category. Moreover, if we consider any difference \( \mu_\ell - \mu_1 = \alpha_\ell \) then we obtain the effect \( \alpha_\ell \) of the \( \ell \)-th category of factor \( A \). Hence, parameter \( \alpha_\ell \) is the expected difference of \( Y \) for an individual belonging in \( \ell \) group (or level) of variable \( A \) in comparison to an individual from the reference group/category of \( A \).

### 5.4.2.1 Sum-to-zero constraints.

According to the name of this parametrization the following constraint is imposed

\[
\sum_{\ell=1}^{L_A} \alpha_\ell = 0. \tag{5.11}
\]

In practice, within the likelihood we substitute one parameter (usually the first or the last one) with the function resulting by the sum-to-zero constraint (5.11). When we substitute the first level then

\[
\alpha_1 = - \sum_{\ell=2}^{L_A} \alpha_\ell. \tag{5.12}
\]

The interpretation of this parametrization is different than the corresponding interpretation of the corner constrained parameters. In STZ, the constant term encapsulates an overall mean effect since

\[
\sum_{\ell=1}^{L_A} \mu'_\ell = L_A \mu_0 + \sum_{\ell=2}^{L_A} \alpha_\ell = L_A \mu_0 \Leftrightarrow \mu_0 = \frac{1}{L_A} \sum_{\ell=1}^{L_A} \mu'_\ell
\]
while parameter $\alpha_\ell$ describes deviations of each level from this overall mean effect. Positive effects induce an increased effect in comparison to the overall mean while negative values induce effects lower than the overall mean level.

### 5.4.3 One way anova model in WinBUGS

In this section, we assume that the data are given by pairs $(a_i, y_i)$ referring to the characteristics of the $i$-th individual. The stochastic part of the likelihood is the same as the one used for normal regression models above. The deterministic part of the likelihood is slightly changed since the mean must be specified as a function of each level of $A$ (see equation 5.10 in page 170). Hence, the likelihood is defined in WinBUGS

```r
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. Thus, for CR parametrization, we set

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

while, for STZ parametrization, we set

```r
```

Alternatively, the group means $\mu'_\ell$ (denoted by $m$ in the WinBUGS code) can be estimated directly. The desired effects can be simply calculated as contrasts of the group means $\mu'_\ell$. Thus we can write

\[
\mu_0 = \mu'_1, \quad \alpha_1 = 0 \quad \text{and} \quad \alpha_\ell = \mu'_\ell - \mu'_1 \quad \text{for} \quad \ell = 2, \ldots, L_A
\]

for CR parametrization and

\[
\mu_0 = \overline{\mu}' = \frac{1}{L_A} \sum_{\ell=1}^{L_A} \mu'_\ell, \quad \text{and} \quad \alpha_\ell = \mu'_\ell - \overline{\mu}' \quad \text{for} \quad \ell = 1, \ldots, L_A
\]

for STZ parametrization. The above model can be specified in WinBUGS using the following syntax

```r
for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- m[a[i]]
}
#
# for corner constraints
mu.cr <- m[1]
alpha.cr[1]<-0
for (l in 2:LA){ alpha.cr[l]<- m[l]-m[1] }
#
# for STZ constraints
mu.stz <- mean(m[])
```
for (l in 1:LA) { alpha.stz[l] <- m[l] - mu.stz }

Finally, we may adopt one parametrization and calculate the parameters of the other one their relation given by

\[
\mu_{STZ}^0 = \frac{1}{L_A} \sum_{\ell=1}^{L_A} \mu'_{\ell} = \frac{1}{L_A} \sum_{\ell=1}^{L_A} (\mu_{0}^{CR} + \alpha_{\ell}^{CR}) = \mu_{0}^{CR} + \frac{1}{L_A} \sum_{\ell=1}^{L_A} (\alpha_{\ell}^{CR})
\]

and

\[
\alpha_{\ell}^{STZ} = \mu'_{\ell} - \mu_{0}^{CR} = \mu_{0}^{CR} + \alpha_{\ell}^{CR} - \mu_{0}^{CR} - \alpha_{\ell}^{CR} = \alpha_{\ell}^{CR} - \alpha_{\ell}^{CR}
\]

for \( \ell = 1, \ldots, L_A \). Thus, in WinBUGS, we can calculate the STZ constraints by the corner constraints model using the code

\# for STZ constraints from CR
mu.stz <- m.cr + mean(alpha.cr[])
for (l in 1:LA) { alpha.stz[l] <- alpha.cr[l] - mean(alpha.aplha[]) }

As a prior for \( \mu \) and \( \alpha_{\ell} \) (for \( \ell = 2, \ldots, L_A \)), we consider a simple normal distribution with mean zero and low precision to express prior ignorance. Hence, the WinBUGS code

\[
\text{mu0 ~ dnorm( 0.0 , 1.0E-4) }
\]

\[
\text{for (l in 2:LA){ alpha[l] ~ dnorm( 0.0, 1.0E-04 )}}
\]

The prior for the precision \( \tau \) is defined as in normal regression models. In the above formulation no prior is imposed on the constrained parameter \( \alpha_1 \) since it is set equal to zero and therefore it does not appear in the likelihood equation. In this case, parameter \( \alpha_1 \) is a constant node while, in STZ parametrization, it is a logical/deterministic node since it is defined as a function of the rest of parameters.

When using the above prior set-up, we must be very careful since, under different parameterizations, we impose different prior distributions on the group means \( \mu'_{\ell} \). In the case that the prior precision is small, differences due to this incompatibility of prior specification will be minor but when prior information is used then these parameters must be specified carefully in order to lead to compatible prior beliefs.

5.4.4 An one way ANOVA example using WinBUGS

\[\text{EXAMPLE 5.2}\]

\[\text{Evaluation of candidate school tutors.}\] The director of a private school wishes to employ a new mathematics tutor. For this reason, the ability of four candidates is examined using the following small study: a group of 25 students was randomly divided in four classes. In all classes, the same mathematical topic was taught for two hours per day for one week. After completing the short course, all students had to undertake the same test. Their grades were recorded and compared; see Table 5.2. The administrator wishes to employ the tutor whose students attained higher performance at the given test.
ANALYSIS OF VARIANCE MODELS

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

Table 5.9. Data for example 5.2 (school tutors’ evaluation data).

Setting up the data and the model code. Data are coded in WinBUGS using two columns/variables: the first one with the student’s grades and the second one corresponding to which candidate tutor was teaching in each student’s class. Within the data we have also defined the number of cases \((n = 25)\) and the number of tutors \((TUTORS = 4)\). Data can be specified using either a list format or a rectangular format; see Table 5.10. for the list format of the data.

```
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, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4))
```

Table 5.10. WinBUGS List format data for example 5.2 (school tutors’ evaluation data).

The model can be defined according to the directions of the previous sub-section. Each group (tutor) mean will be equal to a constant term plus the effect with index specified by the variable `classes`. Initial values will be specified for the constant `mu` and the effect parameter `alpha` except from the first (baseline) level. The code and the initial values is provided in Table 5.11. which follows.

Results. After generating 3000 iterations in total and discarding the initial 1000 iterations, the posterior summaries given in Table 5.12. have been calculated. In this example, we are interested to evaluate the overall performance of each tutor which is encapsulated by each parameter \(\alpha_j\) \((alpha[j])\) for \(j = 1, 2, 3, 4\). Comparing the posterior means and medians of these parameters, we can see that the fourth tutor has higher performance (close to 6 while the rest tutors have negative effects indicating that their performance is below the overall mean). Nevertheless, from the graphical representations of Figures 5.2. we may conclude that the posterior distributions of \(\alpha_j\) are not clearly discriminated, indicating that the between tutors differences are minor. Since the school needs to hire only one tutor, we propose to hire the last one but having in mind that differences in the tutors’ performance in this small study were not indicating clear differences between tutors’ actual abilities.

5.4.5 Two way analysis of variance model

5.4.5.1 The main effects model

We can extend the analysis of variance model to accommodate additional categorical explanatory variables. In the following, we illustrate models with two categorical explana-
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 )

Table 5.11. WinBUGS code and initial values for example 5.2 (School Tutors’ Evaluation); CR parameterization can be fitted by removing the comment sign # in line 10 and adding it to line 8 of the code.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha[1]</td>
<td>-0.5661</td>
<td>8.016</td>
<td>0.1488</td>
<td>-16.27</td>
<td>-0.661</td>
<td>15.43</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>m</td>
<td>68.96</td>
<td>4.561</td>
<td>0.1109</td>
<td>60.0</td>
<td>68.98</td>
<td>78.11</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>s</td>
<td>22.21</td>
<td>3.638</td>
<td>0.08999</td>
<td>16.54</td>
<td>21.74</td>
<td>30.55</td>
<td>1001</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 5.12. Posterior summaries for ANOVA parameters of example 5.2 (School Tutors’ Evaluation).
tory variables (two way analysis of variance models) and then, in the next sub-section, we briefly describe implementation of the general multi-factor ANOVA models.

Let us consider two categorical factors $A$ and $B$ with $L_A$ and $L_B$ levels respectively. Then a natural extension of model (5.8) is to include the additive effect of the additional variable in the expression of the mean $\mu'_{ab}$ for $a$ and $b$ levels of $A$ and $B$ categorical variables respectively. Hence $\mu'_{ab}$ can be written as

$$\mu'_{ab} = \mu_0 + \alpha_a + \beta_b,$$

for $a = 1, 2, \ldots, L_A$ and $b = 1, 2, \ldots, L_B$.

When data are given in tabular form of size $L_A \times L_B$ with $n_{ab}$ observation per cell, then the above model is given by

$$y_{abk} \sim N(\mu'_{ab}, \sigma^2) \text{ and } \mu'_{ab} = \mu_0 + \alpha_a + \beta_b$$

for $k = 1, 2, \ldots, n_{ab}$ and $a, b$ taking values as above.

The above tabular setup is restrictive and difficult to define within WinBUGS model code unless a sample with equal number of observations per cell is considered. A more general setup, equivalent to the formulation for one-way models (see equation 5.10), is given by the following equation

$$y_i \sim N(\mu_i, \sigma^2) \text{ and } \mu_i = \mu_0 + \alpha_{a_i} + \beta_{b_i}$$

for $i = 1, 2, \ldots, n$. In the above model formulation, data are given in the form of $(a_i, b_i, y_i)$ for each subject $i$ with $a_i$ and $b_i$ representing the groups/levels of factors $A$ and $B$ respectively in which $i$ subject belongs.

### 5.4.5.2 Parametrization and parameter interpretation.

Parametrization of the above model can be completed by using STZ or CR constraints for both variables involved in the model. Mixed parametrization (for example CR for one

---

**Figure 5.2.** Posterior Box-plots and error bars for tutors’ effects in example 5.2 (School Tutors’ Evaluation).
variable and STZ for the other) can be used without any difficulty, but the user must be careful with the parameter interpretation. For this reason, such practice is not recommended to users with limited experience in modeling. In the following, we briefly present the parameter interpretation for each parameterization using the same logic as in section 5.4.2.

For the corner constraints, the constant parameter indicates the mean of \( Y \) for the reference categories of both variables \( A \) and \( B \) since \( \mu'_{11} = \mu_0 \) (assuming here the first ones as baseline levels). Effect \( \alpha_a \) indicates the expected group differences between subjects of level \( a \) and subjects belonging in the reference category of \( A \) when both being members of group \( b \) of factor \( B \) since

\[
\mu'_{ab} - \mu'_{1b} = \mu_0 + \alpha_a + \beta_b - \mu_0 - \alpha_1 - \beta_b \\
= \alpha_a - \alpha_1 = \alpha_a .
\]

Similarly, \( \beta_b \) is equal to the expected difference between individuals of level \( b \) and individuals of the reference category of factor \( B \) belonging in the same group \( a \) of variable \( A \) since

\[
\mu'_{ab} - \mu'_{a1} = \mu_0 + \alpha_a + \beta_b - \mu_0 - \alpha_a - \beta_1 \\
= \beta_b - \beta_1 = \beta_b .
\]

For the sum-to-zero constraints, the constant parameter is equal to an overall mean (grand mean) estimate since

\[
\sum_{a=1}^{L_A} \sum_{b=1}^{L_B} \mu_{ab} = \sum_{a=1}^{L_A} \sum_{b=1}^{L_B} (\mu_0 + \alpha_a + \beta_b) \\
= L_AL_B\mu_0 + L_B \sum_{a=1}^{L_A} \alpha_a + L_A \sum_{b=1}^{L_B} \beta_b = L_AL_B\mu_0 \\
\Leftrightarrow \mu_0 = \frac{1}{L_AL_B} \sum_{a=1}^{L_A} \sum_{b=1}^{L_B} \mu_{ab} .
\]

Moreover, since the mean of all effects \( \alpha_a \) is equal to zero, the effect \( \alpha_a \) indicates group differences between level \( a \) and the overall mean effect of \( A \) within each group \( b \) of variable \( B \). Similarly, \( \beta_b \) represents differences between level \( b \) and the overall mean effect of \( B \) within each group \( a \) of variable \( A \).

In the ‘main effects’ model defined above, the effect of the level of each factor is assumed constant across the levels of the second variable. This property is the main characteristic of the ‘main effects’ model and, for this reason, such effects are called additive.

5.4.5.3 The two way interaction model.

Frequently in practice the effect of the explanatory categorical variable \( A \) on a response variable \( Y \) is influenced by another factor \( B \). In such cases, the two factors \( A \) and \( B \) interact concerning their effect on \( Y \) and therefore the additive model is no more valid. A more complicated structure must be defined to encapsulate this ‘interaction’ effect on the mean which can be expressed by

\[
\mu'_{ab} = \mu_0 + \alpha_a + \beta_b + \alpha\beta_{ab}
\]
for \( a = 1, 2, \ldots, L_A \) and \( b = 1, 2, \ldots, L_B \). Parameters \( \alpha \beta_{ab} \) now encapsulate the way that the two categorical factors \( A \) and \( B \) interact and change their effect on \( Y \).

Constraints on interaction term \( \alpha \beta = (\alpha \beta_{ab}) \) must be set in a similar manner as in the main effects. In order to match the \( L_A \times L_B \) group means we need to impose \( L_A + L_B - 1 \) constraints on the interaction parameters \( \alpha \beta \). In the CR parametrization we impose the following constraints

\[
\alpha \beta_{r_a,b} = \alpha \beta_{rb} = 0
\]

for all \( a = 1, 2, \ldots, L_A \) and \( b = 1, 2, \ldots, L_B \); where \( r_a \in \{1, 2, \ldots, L_A\} \) and \( r_b \in \{1, 2, \ldots, L_B\} \) are the reference categories for categorical factors \( A \) and \( B \) respectively. In our case, we consider \( r_a = r_b = 1 \) hence \( \alpha \beta_{11} = 0 \).

Similarly, for the STZ parametrization, we impose constraints

\[
\sum_{a=1}^{L_A} \alpha \beta_{ab} = \sum_{b=1}^{L_B} \alpha \beta_{ab} = 0 . \tag{5.13}
\]

In order to estimate the model, we simply substitute \( L_A + L_B - 1 \) parameters by the corresponding equations resulting from constraints \( (5.13) \). For example, by considering the substitution of the first levels of each factor we end up setting

\[
\alpha \beta_{a1} = - \sum_{b=2}^{L_B} \alpha \beta_{ab} \quad \text{for all} \quad a = 1, 2, \ldots, L_A , \tag{5.14}
\]

\[
\alpha \beta_{1b} = - \sum_{a=2}^{L_A} \alpha \beta_{ab} \quad \text{for all} \quad b = 2, 3, \ldots, L_B . \tag{5.15}
\]

Interpretation is similar to the one described in the previous section but we now have to consider also the interaction term. In corner parameterization \( \mu_0 \) is still the mean of \( Y \) for experimental units in the reference categories of both categorical factors. Effects \( \alpha_a \) now refer to the mean difference between the reference and \( a \) level of variable \( A \) when \( B \) is set to its reference category \( r_a \) since

\[
\mu'_{ar_b} - \mu'_{r_a r_b} = \mu_0 + \alpha_a + \beta_{rb} + \alpha \beta_{ar_b} - \mu_0 - \alpha_{r_a} - \beta_{rb} - \alpha \beta_{r_a r_b} = \mu_0 + \alpha_a - \mu_0 = \alpha_a .
\]

Similarly, \( \beta_b \) refer to the mean difference between the reference and \( b \) level of variable \( B \) when \( A \) is set to its reference category \( r_a \). The interaction term \( \alpha \beta_{ab} \) refers to the additional effect due to the interaction between the two levels since, for \( a, b > 1 \),

\[
\mu'_{ab} - \mu'_{r_a r_b} = \mu_0 + \alpha_a + \beta_b + \alpha \beta_{ab} - \mu_0 - \alpha_{r_a} - \beta_b - \alpha \beta_{r_a r_b} = \mu_0 + \alpha_a + \beta_b + \alpha \beta_{ab} - \mu_0 - \beta_b = \alpha_a + \alpha \beta_{ab} .
\]

The difference \( \mu'_{ab} - \mu'_{r_a r_b} \) is now not the same as the corresponding difference in the reference category (i.e. for \( b = r_b \)). This difference is now affected by the levels of the factor \( B \) which is not the case in the main effects model.

Interpretation of the STZ parameters is equivalent to the corresponding interpretation in CR parametrization but now all comparisons are made with respect to the ‘grand mean’. It is slightly more difficult and, for this reason, corner constraints are frequently used to simplify interpretation.
5.4.5.4 Data in tabular form (equal observations per cell).

Frequently, data for analysis of variance models are provided in a tabular form. In this section we briefly demonstrate the WinBUGS model code when data are provided in such form.

Let us consider two categorical factors $A$ and $B$ with levels $L_A$ and $L_B$. Then the data will be presented in a $L_A \times L_B$ tabular form with $K$ observations in each level combination (cell). Here we assume equal observations per cell. Equivalent approach can be followed in cases of unbalanced data using missing values to fill in empty cells of the defined matrix. When unbalanced data are given in raw - individual type format then the approach described in section 5.4.5.6 which follows is recommended.

Response data $y_{abk}$ refer to the $k$-th observation of $Y$ for the $a$ and $b$ levels of the factors $A$ and $B$ respectively. The structure of the data are depicted in Table 5.13.. We use a $L_A \times L_B \times K$ array to store the data in WinBUGS. In cases where $K = 1$ (i.e. one observation per cell) then we can use a $L_A \times L_B$ matrix.

Table 5.13. Tabular form for two way ANOVA data with $K$ observations per cell.

**Setting up the data.** The simplest approach is to use R to set-up an array of appropriate dimension, then export it to WinBUGS and finally edit the file to meet the requirements of WinBUGS (see section 2.4.6.2 in page 104 for general instructions).

To specify the data in a list format we can use the syntax

```r
list( LA=#LA#, LB=#LB#, K=#K#,
   y = structure(
      .Data=c( y111, y112, ..., y11K,
               y121, y122, ..., y12K, ..., ...,
               y1LB1, y1LB2, ..., y1LBK,
               y211, y212, ..., y21K,
               y221, y222, ..., y22K, ..., ...,
               y2LB1, y2LB2, ..., y2LBK,
               ..., ..., ...,
               yLA11, yLA12, ..., yLA1K,
               yLA21, yLA22, ..., yLA2K, ..., ...,
               yLALB1, yLALB2, ..., yLALBK),
      .Dim = c( #LA#, #LB#, #K# )
   )
)
```
where \( \#L_A \), \( \#L_B \), \( \#K \) are fixed numbers defining the number of levels for each variable and the number of observations per cell. In the above syntax we have specified firstly all observations for cell \((1,1)\) followed by observations for cell \((1,2), \ldots, (1, L_B)\) (levels of factor \(B\) changing first). After data for all categories of \(B\) are defined, then we insert all data for the next level of \(A\) using the same structure etc.

A simpler way is to specify each column \( y[a,b,] \) (or \( y[ ,b,k] \)) separately using the rectangular data format. This is easier to follow although it requires \(L_A \times L_B\) columns to be specified.

**The model code.** Having specified the data, the model likelihood is specified using a triple loop in order to define all elements of the response array \(y\). Thus the main effects model is defined by

\[
y_{abk} \sim N(\mu_{abk}^{\prime}, \tau^{-1}) \quad \text{and} \quad \mu_{abk}^{\prime} = \mu_0 + \alpha_a + \beta_b
\]

for \(a = 1, 2, \ldots, L_A\), \(b = 1, 2, \ldots, L_B\) and \(k = 1, 2, \ldots, K\), while the corresponding WinBUGS code is given by

```r
for (a in 1:LA){
  for (b in 1:LB){
    for (k in 1:K){
      y[a,b,k] ~ dnorm( mu[a,b,k], tau )
      mu[a,b,k] <- mu0 + alpha[a] + beta[b]
    }
  }
}
```

Priors and constraints are defined as in the one way ANOVA model.

For the interaction model, we only need to change the specification of the mean since

\[
\mu_{ab}^{\prime} = \mu_0 + \alpha_a + \beta_b + \alpha\beta_{ab}
\]

Thus, the fifth line of the code above is substituted by the following command

```r
mu[a,b,k] <- mu0 + alpha[a] + beta[b] + ab[a,b]
```

Note that the interaction term \(ab\) is defined as a matrix of dimension \(L_A \times L_B\). Priors must be defined for all components of \(\alpha\beta\) except for the ones referring to the baseline levels \(r_a\) and \(r_b\).

For the corner constraints, with the first level as baseline for both variables, we use the following syntax to specify the priors and the constraints

```r
# corner constraints for baseline levels (1st levels)
ab[1,1] <- 0
for (a in 2:LA){ ab[a,1]<-0 }
for (b in 2:LB){ ab[1,b]<-0 }
# prior distributions for the rest interaction parameters
for (a in 2:LA){
  for (b in 2:LB){
    ab[a,b]~dnorm(0.0, 1.0E-04)
  }
}
```

STZ constraints will be defined by substituting the first four lines of the above code with the following syntax which represent equations 5.15 and 5.14.

```r
# corner constraints for baseline levels (1st levels)
for (a in 1:LA){ ab[a,1]<- -sum( ab[a,2:LB] ) }
for (b in 2:LB){ ab[1,b]<- -sum( ab[2:LA,b] ) }
```
5.4.5.5 A two way ANOVA example.

- EXAMPLE 5.3

**Schizotypal Personality Data.** Let us consider the data presented in Table 5.14, inspired by a student survey examining the association between schizotypal traits and impulsive and compulsive buying behavior of University students (Iliopoulou 2004). Cell values of the Table represent scores of a psychometric scale called schizotypal personality questionnaire (SPQ, Raine 1991) of individuals which are close to the median values of each group. In the original study of Iliopoulou (2004) no female students leaving in villages reported high Economic status (indicated by NA). In chapter 8 we illustrate how we can deal with missing response values using the Bayesian approach. Nevertheless, here, we have substituted the missing values with two imaginary values (indicated within brackets) for illustration purposes. In this example we are interested in the effect of the student’s

1. family economic status (categorical with three levels: low, medium, high),
2. city residence factor (binary: yes, no) and
3. gender

don the SPQ total score (Raine 1991).

<table>
<thead>
<tr>
<th>Economic Status</th>
<th>Gender</th>
<th>Area of Residence</th>
<th>City</th>
<th>Village</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (1)</td>
<td>Male</td>
<td>9</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Medium (2)</td>
<td>Female</td>
<td>25</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>High (3)</td>
<td>Male</td>
<td>23</td>
<td>12</td>
<td>NA (24)</td>
</tr>
</tbody>
</table>

**Table 5.14.** Data for example 5.3 (Schizotypal personality data).

Raine’s SPQ scale is a 74-item self-administered questionnaire used to measure the concepts related to schizotypal personality. The questionnaire consists of binary zero-one (yes-no) items. It provides subscales for nine schizotypal features as well as an overall scale for schizotypy. These nine specific characteristics of a ‘schizotypal personality’ are defined in the DSM-III-R diagnostic and statistical manual of mental disorders edited by the American Psychiatric Association (1987) and are measured by an SPQ subscale calculated as the sum of the questionnaire items that refer to each schizotypal characteristic.

A ‘schizotypal personality’ suffers from minor episodes of “pseudoneurotic” problems. In general, the prevalence rate of schizotypy is about 10% in the general population. The importance of schizotypal personality in psychiatric research is prominent for two reasons: Firstly, shizotypal subjects have increased risk to develop schizophrenia during their life. Secondly, since they are healthy persons, they can participate in psychiatric/psychological research studies (by completing questionnaires - psychometric instruments) in which schizophrenia cases are unable to do. In this example we focus on the total SPQ score.
Two way interaction model. In this subsection, we ignore the “city residence factor” assuming that we have two observations for each combination of Economic status (factor $A$) and gender (factor $B$). Hence we need to setup the data in $3 \times 2 \times 2$ array ($L_A = 3$, $L_B = 2$ and $K = 2$).

Following the directions above we firstly rearrange the data in the following way

<table>
<thead>
<tr>
<th>Economic Status</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (1)</td>
<td>9</td>
<td>14</td>
<td>Male</td>
</tr>
<tr>
<td>Medium (2)</td>
<td>25</td>
<td>26</td>
<td>Male</td>
</tr>
<tr>
<td>High (3)</td>
<td>23</td>
<td>24</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>22</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>25</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13</td>
<td>Female</td>
</tr>
</tbody>
</table>

We can now define the data in WinBUGS using the following syntax

```r
list(LA=3, LB=2, K=2,
    y = structure(  
        .Data=c( 9, 14, 18, 29,
                25, 26, 22, 25,
                23, 24, 12, 13),  
        .Dim = c( 3, 2, 2 )  
    )  
)
```

To ensure that the data have been imported correctly in WinBUGS, compile the model and go to “Info>Node Info...”, insert “y” to view the response data and press the “Values” box. The values of “y” will be printed in the log file for cross checking. The correct form in this example is the following

```
y[1,1,1] 9.0
y[1,1,2] 14.0
y[1,2,1] 18.0
y[1,2,2] 29.0
y[2,1,1] 25.0
y[2,1,2] 26.0
y[2,2,1] 22.0
y[2,2,2] 25.0
y[3,1,1] 23.0
y[3,1,2] 24.0
y[3,2,1] 12.0
y[3,2,2] 13.0
```

Alternatively, we can use four columns of length equal to 3 to setup the data using the following rectangular data format

```r
list(LA=3, LB=2, K=2)
y[,1,1] y[,1,2] y[,2,1] y[,2,2]  
9 14 18 29
25 26 22 25
23 24 12 13
```
In the list data format of the above syntax we define the dimensions of the table needed for the specification of the likelihood in the model code. The full code for the interaction model is given in Table 5.15. Note that instead of alpha, beta, ab we have used the names econ, gender, econ.gender to reflect the actual names of the factors.

```plaintext
model{
    # model's likelihood
    for (a in 1:LA){
        for (b in 1:LB){
            for (k in 1:K){
                y[a,b,k] ~ dnorm( mu[a,b,k], tau )
                mu[a,b,k] <- mu0 + econ[a] + area[b] + econ.area[a,b]
            }
        }
    }

    #### CR Constraints
    econ[1] <- 0.0
    area[1] <- 0.0
    econ.area[1,1] <- 0.0
    for (a in 2:LA){econ.area[a,1]<-0.0}
    for (b in 2:LB){econ.area[1,b]<-0.0}

    # priors
    mu0~dnorm( 0.0, 1.0E-04)
    for (a in 2:LA){econ[a]~dnorm( 0.0, 1.0E-04)}
    for (b in 2:LB){area[b]~dnorm( 0.0, 1.0E-04)}
    for (a in 2:LA){
        for (b in 2:LB){
            econ.area[a,b]~dnorm( 0.0, 1.0E-04)
        }
    }
    tau ~dgamma( 0.01, 0.01)
    s <- sqrt(1/tau) # precision
}

INITS
list( mu0=1.0, econ=c(NA, 0,0), area=c(NA, 0),
    econ.area=structure(.Data=c(NA, NA, NA, 0, NA, 0), .Dim=c(3,2)),
    tau=1.0 )
```

Table 5.15. WinBUGS code and initial values for example 5.3 (Schizotypal personality data).

Results. Results obtained after 3000 iterations (discarding the initial 1000 iterations as a burnin) are provided in Table 5.16. and Figure 5.3. From the box-plots we can observe that

- Both interaction terms are far away from zero indicating that this term must be included in the model.
• The 95% posterior intervals of the two interaction parameters (econ.gender_{22} and econ.gender_{32}) have common values indicating that they can be considered as a-posteriori equal.

• Both gender and economic status effects are far away from zero indicating that they influence the mean SPQ score.

• Small differences are observed between the posterior distributions of the medium and high economic status effects.

To be more specific, the mean SPQ score for a male student with low economic family status is a-posteriori expected to be equal to 11.5. Females students with the same economic status are a-posteriori expected to score about 12 units higher (i.e. 23.5). Moreover, we observe a positive effect of economic status on male students since students with medium family economic status are a-posteriori expected to score 14 points additional to the ones scored by male students with low economic status. The corresponding increase is lower for the high economic group (about 12 points) indicating no separation in the effects of these two groups.

The effect of economic status is opposite for female students since the interaction term indicates that increase in the economic status of female student from low to medium does not change the a-posteriori expected mean SPQ score (since econ[2]+econ.gender[2,2] = 14.09 − 14.19 = −0.1). Moreover, a female student with high economic family status is a-posteriori expected to get an SPQ score of 11 units lower than a female from low economic status (since econ[3]+econ.gender[3,2] = 12.01 − 23.08 = −11.07).

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>econ.gender[3,2]</td>
<td>−23.08</td>
<td>6.095</td>
<td>0.375</td>
<td>−36.51</td>
<td>−22.83</td>
<td>−11.74</td>
<td>1001 2000</td>
</tr>
<tr>
<td>mu0</td>
<td>11.46</td>
<td>3.012</td>
<td>0.1542</td>
<td>5.57</td>
<td>11.43</td>
<td>17.62</td>
<td>1001 2000</td>
</tr>
<tr>
<td>s</td>
<td>4.081</td>
<td>1.4</td>
<td>0.06154</td>
<td>2.3</td>
<td>3.793</td>
<td>7.522</td>
<td>1001 2000</td>
</tr>
</tbody>
</table>

Table 5.16. Posterior summaries for the parameters of the interaction two ANOVA model of example 5.3 (Schizotypal Personality Data).

The above comments can be depicted in an interaction plot based on the posterior means of expected SPQ scores (i.e. μ’s) for each combination of the categories of the two variables under consideration. In Figure 5.4, we have used one line for each gender. Note that when the main effects model is used then these two lines will not cross while the lines connecting means of different levels will be parallel indicating equal effect of the categorical variable across all levels of the other other one. From this plot it is clear that the economic status is positively associated with the mean SPQ score for males and negatively associated for females. Moreover, the difference between males and females is minor for the medium economic group. Finally, the mean SPQ score is a-posteriori expected to be equal for males with low economic status and females with high economic status.

In Figure 5.4.(a) the original interaction plot is provided while in Figure 5.4.(b) the same plot is annotated with notes concerning each parameter. The curly brackets refer to the interaction terms which indicate deviations from the main effects model. Hence the expected
SPQ difference between females and male students for medium and high economic status is given from the differences $\text{econ.gender}_{22} - \text{gender}_2$ and $\text{econ.gender}_{32} - \text{gender}_2$ respectively. In order to be more informative, the above plot can be enriched using posterior box-plots or error bars for each group; see, for example, in Figure 5.5.

**Figure 5.4.** Interaction plot based on posterior means of expected SPQ score in example 5.3 (Schizotypal personality data).

### 5.4.5.6 Individual type data.

Individual type data are usually available in practice. In this format each variable data are provided in a tabular form with $n$ rows and $p + 1$ columns. Each row corresponds to one individual while each column in one variable (as usually in statistical packages). Using this approach we can easily fit an ANOVA model for unbalanced data. The main effects model can be easily programed using the syntax

```r
for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- mu0 + alpha[ a[i] ] + beta[ b[i] ]
}
```
Figure 5.5. Interaction plot based on 95% posterior intervals of expected SPQ score in example 5.3 (Schizotypal personality data).
while for the interaction term we only substitute the expression for $\mu$ by

$$
\text{mu}[i] \leftarrow \mu_0 + \alpha[a[i]] + \beta[b[i]] + \text{ab}\{a[i], b[i]\}
$$

in order to also accommodate the interaction term $\text{ab}$.

For the example 5.3 the data can inserted using the following syntax

```r
list( n=12, LA=3, LB=2,
      g=c(1,1,1,2,2,2,1,1,1,2,2,2),
      e=c(1,2,3,1,2,3,1,2,3,1,2,3),
      y=c(9, 25, 23, 18, 22, 12, 14, 26, 24, 29, 25, 13) )
```

in list format, or by

```r
list(n=12, LA=3, LB=2)
  g[ ] e[ ] y[ ]
1 1 1 9
1 2 2 25
1 3 3 23
2 1 1 18
2 2 2 22
2 3 3 12
1 1 1 14
1 2 2 26
1 3 3 24
2 1 1 29
2 2 2 25
2 3 3 13
END
```

when using the rectangular data format. In the above data format, $g[ ]$ and $e[ ]$ denote the gender and economic status of each student of the data. The expression for the mean $\mu_i$ will be given by

$$
\text{mu}[i] \leftarrow \mu_0 + \text{gender}[g[i]] + \text{econ}[e[i]]
$$

for the main effects model and by

$$
\text{mu}[i] \leftarrow \mu_0 + \text{gender}[g[i]] + \text{econ}[e[i]] + \text{econ.gender}[e[i], g[i]]
$$

for the interaction model.

Note that, since by this way we can easily handle unbalanced data, for example 5.3 we can eliminate the two observations which correspond to missing values at the original dataset and rerun the MCMC algorithm. Table 5.17. presents results after removing these two values. Results are slightly different but the main conclusions remain the same as in the full data analysis.

The “individual” type data format is recommended since it is more general and can be implemented in most models by slightly modifying the above model code. Moreover, unbalanced data can be easily fitted using this approach.
Table 5.17. Posterior summaries for the parameters of the interaction two ANOVA model of example 5.3 (Schizotypal Personality Data) using unbalanced data.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu₀</td>
<td>11.44</td>
<td>4.507</td>
<td>0.2476</td>
<td>3.277</td>
<td>11.41</td>
<td>19.8</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>s</td>
<td>5.473</td>
<td>3.123</td>
<td>0.222</td>
<td>2.583</td>
<td>4.741</td>
<td>12.43</td>
<td>1001</td>
<td>2000</td>
</tr>
</tbody>
</table>

5.4.5.7 Multifactor Analysis of Variance.

ANOVA models can be extended to accommodate more than two variables. The main effects model includes no interaction and has similar interpretation as the one described for the two way model. Including higher order interaction complicates the model and the user must be careful when using them.

Although we can extend the above approach for multifactor ANOVA models, we will face the problem that the imposed constraints result in a complicated model structure which needs a considerable amount of code in order to be fitted in WinBUGS. Such models can be fitted in a more straightforward manner using dummy variables which are described in detail in section 5.5.2 which follows.

5.5 INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

In this section incorporation of the categorical covariates in the usual regression models is presented. The use of dummy variables for the two parameterizations described in the previous section is firstly illustrated. Then focus is given on ANOVA models using dummy variables and on the analysis of covariance (ANCOVA) models in which both qualitative and quantitative variables are used as covariates.

5.5.1 Dummy variables and design matrices.

Categorical variables can be incorporated in regression models via the use of dummy variables. These dummy variables identify which parameters must be added (and how) to the linear predictor.

Let us consider the simple one-way ANOVA with a categorical factor $A$ and $L_A$ categories. Under the CR constraints the mean of an individual observation $i$ with data $(y_i, a_i)$ is given by

$$
\mu_i = \mu_0 \quad \text{when} \quad a_i = 1
$$

$$
\mu_i = \mu_0 + \alpha_{a_i} \quad \text{when} \quad a_i > 1.
$$

The aim is to express the mean $\mu_i$ as a linear combination of the covariates equivalent to (5.2). Thus, we can write

$$
\mu_i = \mu_0 + \alpha_2 D_{a_2}^A + \alpha_3 D_{a_3}^A + \ldots + \alpha_{L_A} D_{a_{L_A}}^A
$$

where $D_{a_\ell}^A$, $\ell = 1, 2, \ldots, L_A$ are dummy variables defined as

$$
D_{a_\ell}^A = 1 \text{ if } a_\ell = \ell \text{ and } D_{a_\ell}^A = 0 \text{ otherwise.}
$$

(5.16)
As we can observe, parameters \((\mu_0, \alpha_2, \ldots, \alpha_L)\) play the same role as parameters \(\beta = (\beta_0, \beta_1, \ldots, \beta_p)\) in usual regression models. In all cases the number of dummies that we need to use in order to define a model will be equal to \(L - 1\); where \(L\) are the number of level of the variable under consideration.

Similarly, for the STZ parametrization the mean of an individual observation \(i\) with data \((y_i, a_i)\) is given by

\[
\begin{align*}
\mu_i &= \mu_0 - \alpha_2 - \alpha_3 \ldots - \alpha_L \quad \text{when } a_i = 1 \\
\mu_i &= \mu_0 + \alpha_{a_i} \quad \text{when } a_i > 1.
\end{align*}
\]

The dummy variables here are slightly more complicated since

\[
\begin{align*}
D_{A,\text{stz}}^{A,1} &= 1 \quad \text{if } a_i = \ell \\
D_{A,\text{stz}}^{A,1} &= -1 \quad \text{if } a_i = 1 \\
D_{A,\text{stz}}^{A,1} &= 0 \quad \text{otherwise.}
\end{align*}
\] (5.17)

In the above, we have expressed the first category as a function of the rest of the parameters. If we wish to omit from the parameters a different category, then the dummy variables must be modified accordingly. Note that the STZ dummy variables can be easily expressed as the difference of the CR dummy variables. Hence, assuming that the first category will serve as baseline, we can write

\[
D_{A,\text{stz}}^{A,1} = D_{A,1}^A - D_{1}^A.
\] (5.18)

The use of dummy variables simplifies the model specification since its structure is the same whatever type of parametrization or data we use. Moreover, interaction terms are directly defined by the products of the corresponding dummy variables. For example, the product \(D_{A}^{A} D_{B}^{B}\) will provide the parameter \(\alpha_{1}\beta_{1}\) of the interaction between factors \(A\) and \(B\) for the corner constraints. Hence we can write

\[
D_{iab}^{AB} = D_{ia}^{A} D_{ib}^{B}
\] (5.19)

where \(D_{iab}^{AB}\) is the dummy variables for the interaction between \(a\) and \(b\) levels of factors \(A\) and \(B\) for \(i\) subject. This multiplicative property is also true for higher interaction terms and is convenient in terms of WinBUGS programming since the definition of the constraints (as in section 5.4) is avoided. Only actually used parameters are defined in the model while omitted parameters (as in the STZ parametrization) can be directly monitored using a simple deterministic/logical node in WinBUGS model code.

Dummy variables can be also specified within WinBUGS using the following commands. For a categorical variable \(A\) with \(L_A\) levels in which \(i\) subject belongs in \(a_i\) category we can specify the dummy variables for CR parametrization using the following syntax

```r
for (i in 1:n){
  D.A2[i] <- equals(a[i],2)
  ....
  D.ALA[i] <- equals(a[i],LA)
}
```

where \(LA\) is the number of levels of factor \(A, L_A\). The above can be additionally simplified is we use a matrix of dimension \(n \times L_A\) to specify the dummy variables for factor \(A\). Then we can write

```r
for (i in 1:n){
  D.A[i] <- matrix(a[i], nrow=n, ncol=L_A)
}
```
for (l in 1:LA){
    D.A[i,l] <- equals(a[i],l)
}}

Using the above syntax we set \( D.A_{i\ell} = 1 \) if \( a_i = \ell \) and zero otherwise following the definition of the dummy variables given in (5.16). Moreover, we have specified \( L_A \) dummy variables instead of the required \( L_A - 1 \). In the linear predictor (equation for \( \mu_i \)) we must use only \( L_A - 1 \) of the columns of matrix \( D.A \). The column excluded by the linear predictor corresponds to the baseline (or reference) category.

Similarly, following (5.17), the syntax

for (i in 1:n){
    for (l in 1:LA){
        DSTZ.A[i,k] <- equals(a[i],l) - equals(a[i],1)
    }
}}

produces the corresponding dummies for the STZ parametrization. The difference between STZ and CR dummy variables is in the first (baseline) category. In STZ parametrization, all dummies are set equal to minus one \((-1)\) for the reference category \( DSTZ.A_{i\ell} = -1 \) for all \( \ell = 2, \ldots, L_A \) instead of zero in the CR parametrization. Here, the first column must be omitted from the specification of \( \mu_i \). Changing the reference category also affects values for all dummy variables. Alternatively, the STZ dummy variable can be defined by (5.18) and, thus, specify them in WinBUGS by

\[
\]

following the syntax used for the CR constraints. Alternatively, dummy variables can be defined in another statistical package and then imported in WinBUGS model code as ready-to-use data.

Matrix \( X \) of dimension \( n \times p \) with columns the constant term, the dummy variables of the factors and their corresponding interactions involved in the model is called design matrix. When mixed (dummy and continuous) type of variables are included in the linear predictor then it is called data matrix as in the regression model. We can directly use matrix \( X \) to specify our model as described in section 5.3.4 for the regression model. Such strategy will considerably simplify the model specification. The design matrix can be easily constructed either within WinBUGS model code (following similar approach as for the construction of dummy variables) or outside WinBUGS and then import the design matrix in the data of the WinBUGS code; for a detailed example see section 5.5.2 which follows.

Since matrix \( X \) can be defined whatever the type of variables we use, all prior distributions described in section 5.3.2 can be used without any difficulty.

Moreover, the algebra related to dummy variables is intriguing since, in cases of balanced data, all design matrices can be calculated using Kronecker products. More specifically, the STZ parametrization has interesting properties resulting in independent posterior distributions of parameters related to different model terms (main effects and interaction parameters); see Ntzoufras (1999) for a brief discussion.

5.5.2 Analysis of variance models using dummy variables.

A three way ANOVA model for schizotypal personality data (Example 5.17.).
In the following we directly define our WinBUGS model using dummy variables for the unbalanced dataset. The additional variable city, ignored in the two way approach, is also included in the data. Dummy variables for both parameterizations are given in Table 5.3. Code for the full three way model is provided in Table 5.19.

<table>
<thead>
<tr>
<th>Economic Gender</th>
<th>Corner Constraints</th>
<th>STZ Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>City</td>
<td>$D^\text{gender}_2$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.18. Dummy variables for example 5.3.

To avoid definition of each parameter in separate scalar nodes, the design matrix $X$ as well as a vector node for all parameters ($\beta$) may be used instead. This approach requires a shorter code which is presented in Table 5.20. The user must identify the location in which each parameter is stored in the vector node $\beta$. Specific parameters can be additionally stored in a separate deterministic node vector in order to monitor only parameters of interest (for example the three way interactions).

Note that in the above model, we have 10 observations, and 13 parameters. This model is clearly over-parametrized given the available set of data. Although in classical modeling approach, such model cannot be fitted, within the Bayesian context the posterior distribution can be estimated for all parameters. The difference is that for the non-identifiable parameters their posterior coincides with their corresponding prior distribution. This implies that no additional information can be extracted from the data for the estimation of these parameters and therefore they can be assumed as redundant.

In the following, we start by fitting the full three way interaction model and then remove the terms that are not important since their posterior distributions are scattered around zero. Starting from the the three way model, we observe that the two parameters of three way interaction term lie around zero with large variance indicating that the likelihood that we may remove the parameters of this term.

Table 5.21. provides posterior summaries for all parameters using the matrix-vector code. In the last column we provide the corresponding parameter name if we use the first (simpler) programming approach. By this table we observe two major points: Firstly, both interaction terms are centered around zero, hence we may remove them. Secondly, parameters $\beta[12]$ and $\beta[10]$ (terms gender.econ.city_{232} and econ.city_{32} respectively) have posterior summaries that match with the corresponding summaries of the $N(0,10^4)$ prior distribution. This is a clear indication that the data do not contribute any information to the posterior distributions of these two parameters.
INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

model{
  for (i in 1:n){
    # CR dummy variables
    D.gender[i]<-equals(g[i],2)
    D.econ2[i]<-equals(e[i],2)
    D.econ3[i]<-equals(e[i],3)
    D.city[i]<-equals(ci[i],2)
    # STZ dummy variables
    DStz.gender[i]<-equals(g[i],2)-equals(g[i],1)
    DStz.econ2[i]<-equals(e[i],2)-equals(e[i],1)
    DStz.econ3[i]<-equals(e[i],3)-equals(e[i],1)
    DStz.city[i]<-equals(ci[i],2)-equals(ci[i],1)
  }
  # model's likelihood
  for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- mu0 + gender * D.gender[i] + econ[2]*D.econ2[i]
    + econ[3]*D.econ3[i] + city*D.city[i]
    + gender.econ[2]*D.gender[i]*D.econ2[i]
    + gender.econ[3]*D.gender[i]*D.econ3[i]
    + gender.city*D.gender[i]*D.city[i]
    + econ.city[2]*D.econ2[i]*D.city[i]
    + econ.city[3]*D.econ3[i]*D.city[i]
    + gender.econ.city[2]*D.gender[i]*D.econ2[i]*D.city[i]
    + gender.econ.city[3]*D.gender[i]*D.econ3[i]*D.city[i]
  }
  # set zero all nuisance parameters
  econ[1] <- 0.0
  econ.city[1]<-0.0
  gender.econ[1]<-0.0
  gender.econ.city[1]<-0.0
  # priors
  mu0~dnorm( 0.0, 1.0E-04)
  gender ~ dnorm( 0.0, 1.0E-04)
  city ~ dnorm(0.0, 1.0E-04)
  gender.city ~ dnorm(0.0, 1.0E-04)
  for (k in 2:3){
    econ[k] ~ dnorm( 0.0, 1.0E-04)
    gender.econ[k] ~ dnorm( 0.0, 1.0E-04)
    econ.city[k] ~ dnorm( 0.0, 1.0E-04)
    gender.econ.city[k] ~ dnorm( 0.0, 1.0E-04)
  }
  tau ~dgamma( 0.01, 0.01)
  s <- sqrt(1/tau) # precision
}

INITS
list( mu0=1.0, econ=c(NA, 0, 0), gender=0, city=0, gender.city=0, gender.econ=c(NA,0,0),
      econ.city=c(NA,0,0), gender.econ.city=c(NA,0,0), tau=1.0)

DATA (LIST)
list(n=10)
g[] e[] ci[] y[]
  1 1 9 1 1 25
  1 2 1 2 1 23
  1 1 1 3 1 18
  2 2 1 2 1 22
  2 3 1 2 1 12
  1 1 2 1 1 14
  1 2 2 2 1 26
  2 1 2 2 1 29
  2 2 2 2 1 25
END

Table 5.19. WinBUGS code for the full three way model for example 5.3 (Schizotypal
Personality Data) using unbalanced data.
model{
  # Creating the design matrix X
  for (i in 1:n){
    X[i,1]<- 1.0 # constant term
    # CR dummy variables
    X[i,2]<-equals(g[i],2) # gender
    X[i,3]<-equals(e[i],2) # econ2
    X[i,4]<-equals(e[i],3) # econ3
    X[i,5]<-equals(ci[i],2) # city
    # STZ dummy variables
    # X[i,2]<-equals(g[i],2)-equals(g[i],1)
    # X[i,3]<-equals(e[i],2)-equals(e[i],1)
    # X[i,4]<-equals(e[i],3)-equals(e[i],1)
    # X[i,5]<-equals(ci[i],2)-equals(ci[i],1)
    # specification of interaction terms
    X[i,6]<-X[i,2]*X[i,3] # gender*econ2
    X[i,7]<-X[i,2]*X[i,4] # gender*econ3
    X[i,8]<-X[i,2]*X[i,5] # gender*city
    X[i,9]<-X[i,3]*X[i,5] # econ.city2
    X[i,10]<-X[i,4]*X[i,3] # econ.city3
    X[i,11]<-X[i,2]*X[i,3]*X[i,5] # gender.econ.city2
    X[i,12]<-X[i,2]*X[i,4]*X[i,5] # gender.econ.city3
  }
  # model’s likelihood
  for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- inprod( X[i,], beta[] )
  }
  # priors
  for (j in 1:12){ beta[j] ~dnorm( 0.0, 1.0E-04) }
  tau ~dgamma( 0.01, 0.01)
  s <- sqrt(1/tau) # precision
}

INITS
list( beta=c(0,0,0,0,0,0,0,0,0,0,0,0), tau=1.0)

Table 5.20. WinBUGS code for the full three way model for example 5.3 using design matrix X and parameter vector beta (Schizotypal Personality Data); data are the same as in Table 5.19.
In the second step, we remove the three way interaction term (by eliminating $X_{i,11}$ and $X_{i,12}$ and $\beta_{11:12}$ from the model code) and rerun the algorithm. Results are similar as above and therefore we proceed by further removing the interaction term $econ \cdot city$. We proceed by successively fitting models $Gender \cdot Econ$ $+$ $Gender \cdot City$, $Gender \cdot Econ$ $+$ $City$ and $Gender \cdot Econ$. In the above notation the asterisk indicates that all lower interaction terms (and effects) are also included in the model. For example $Gender \cdot Econ$ means that the main effects $Gender$ and $Econ$ are also included in the model. Box plots of the fitted models are provided in Figure 5.6. From this box plot we observe that both models $Gender \cdot Econ$ $+$ $City$ and $Gender \cdot Econ$ seem to be satisfactory in the sense that the zero value lies at the tail areas for all parameters. Here we adopt model $Gender \cdot Econ$ $+$ $City$ since only for one parameter the zero value lies within the 95% posterior interval and this is the lower boundary of this interval. Note that model $Gender \cdot Econ$ is the same as in the two way analysis but here we have considered the unbalanced data (10 observations instead of 12). Formal model comparisons follow in chapters 8 and 9.

In the above strategy we start from the full model including all higher order interaction terms and proceed by removing firstly higher order interaction terms. We do not remove lower order interaction terms if these are nested to higher order one (hierarchically structured models). This approach leads to models that are easier to interpret and is made only for convenience. The user may remove main effects or lower order interaction terms but he must be very careful with interpretation. This approach it is not recommended unless the problem setup forces us to follow such strategy.

Posterior summaries of the suggested model are provided in Table 5.22. Interpretation of the parameters is similar as in the two way analysis. Additionally, in this model, city resident students are a posteriori expected to score five SPQ points ($\beta_5 = 5$) higher than students leaving in rural areas of the same gender and economic status.

### 5.5.3 Analysis of covariance models

Using both qualitative and quantitative variables as explanatory variables results to ‘analysis of covariance’ (ANCOVA) models. These models are the natural extension of regression
NORMAL REGRESSION MODELS

Figure 5.6. Box plots of parameters for full three way interaction model for example 5.3 (Schizotypal personality data); Asterisk indicates that all lower interaction terms are also included in the model.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error 2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start term</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta[1]</td>
<td>9.014</td>
<td>3.524</td>
<td>0.07379</td>
<td>9.075</td>
<td></td>
<td>1001 (mu0)</td>
</tr>
<tr>
<td>beta[2]</td>
<td>11.91</td>
<td>4.604</td>
<td>0.09902</td>
<td>11.96</td>
<td>20.76</td>
<td>1001 (gender)</td>
</tr>
<tr>
<td>beta[4]</td>
<td>14.06</td>
<td>5.673</td>
<td>0.127</td>
<td>2.833</td>
<td>14.1</td>
<td>1001 (econ3)</td>
</tr>
<tr>
<td>beta[5]</td>
<td>5.079</td>
<td>3.297</td>
<td>0.07379</td>
<td>-1.631</td>
<td>5.001</td>
<td>1001 (city)</td>
</tr>
<tr>
<td>beta[7]</td>
<td>-23.0</td>
<td>7.878</td>
<td>0.1796</td>
<td>-37.84</td>
<td>-23.25</td>
<td>1001 (gender.econ3)</td>
</tr>
<tr>
<td>s</td>
<td>3.979</td>
<td>2.298</td>
<td>0.1045</td>
<td>1.741</td>
<td>3.377</td>
<td>1001 (st.dev.)</td>
</tr>
</tbody>
</table>

Table 5.22. Posterior summaries for the parameters of model Gender * Econ + City of example 5.3 (Schizotypal Personality Data) using unbalanced data; 2000 MCMC iterations were used.
INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

and analysis of variance models. They initially were used as an extension of ANOVA models in order to test for differences in the means of the response variable after adjusting for the effect of one or more numerical variables. By this way, variation which can be predicted by other numerical covariates is removed before comparing differences between groups. In practice such model do not demonstrate any difficulties in terms of estimation or model building. The interesting part is the interpretation of the models and their structure when we incorporate interaction terms between categorical and numerical variables. Such interactions control the slopes of the regression lines for different groups. Two are the most important models: ‘the parallel lines’ and ‘different slopes’ models.

In order to explain briefly the resulted ANCOVA models, we will describe in more detail the simpler case with one quantitative and one qualitative variable. Extensions when using more explanatory variables can be done in a similar manner.

5.5.3.1 Models using one quantitative and one qualitative variable.

Let us assume that we have available data \((X_i, A_i, Y_i)\) with \(X_i, A_i\) representing the numerical and categorical explanatory variables respectively and \(Y_i\) is the continuous response variable for \(i\) subject. Then the following models can be fitted

1. Constant model: \(\mu_i = \beta_0\),
2. Common line for all groups: \(\mu_i = \beta_0 + \beta_1 X_i\),
3. Constant mean within each group: \(\mu_i = \beta_0 + \alpha_{a_i}\),
4. Parallel line (or common slope) model: \(\mu_i = \beta_0 + \beta_1 X_i + \alpha_{a_i}\),
5. Common intercept model: \(\mu_i = \beta_0 + \beta_1 X_i + \delta_{a_i} X_i\),
6. Separate regression lines for each group: \(\mu_i = \beta_0 + \beta_1 X_i + \alpha_{a_i} + \delta_{a_i} X_i\).

Models 1-3 are already known from simple regression and ANOVA models. More specifically the first model is the simple constant model ‘estimating’ the grand mean of the sample. The second model is a simple regression model since it ignores the grouping variable \(A\) and fits a single regression line to the data. Finally, the third model is a usual ANOVA model which estimates the means for each category of \(A\) by ignoring the numerical variable \(X\). Here we focus our attention in the last three models which are the actual ANCOVA models. The fourth model assumes that the effect of each explanatory variables is additive and is not influenced by the other one while the last one assumes that the effect of the numerical variable if different for each group of the categorical variable. When considering ANCOVA models, we may also construct common or different intercept models by making assumptions about parameters \(\alpha_\ell\) resulting to the fifth model of the list. More details concerning this aspect are provided in the paragraphs which follow. Graphical representation of these models is provided in Figure 5.7.

5.5.3.2 The parallel lines model.

The parallel lines model can be summarized using the following equations

\[
Y_i \sim \text{Normal}(\mu_i, \sigma^2) \\
\mu_i = \beta_0 + \beta_1 X_i + \alpha_{a_i} \quad \text{for} \quad i = 1, \ldots, n.
\] (5.20)
Figure 5.7. Illustration of the possible models with one quantitative and one qualitative explanatory variable.
The mean of the response variable under this model can be effectively summarized by

$$\mu_i = \beta_{0i} + \beta_1 X_i$$  \hspace{1cm} (5.21)

where $\beta_{0i} = \beta_0 + \alpha_i$. From the above it is clear that the regression lines for each group of factor $A$ will have different intercepts and share a common slope (i.e. will be parallel). Such models are also termed as common slope models.

As in ANOVA models we need to impose one constraint on the parameters $\alpha_i (i = 1, 2, \ldots, L_A)$ of the categorical variable. We can use either CR or STZ constraints depending on the parameter interpretation we prefer. Hence the model parameters involved in the linear predictor under this model will be

$$\beta = (\beta_0, \beta_1, \alpha_2, \ldots, \alpha_{L_A})$$

omitting the $\alpha_1$ as the parameter corresponding to the baseline or the reference category.

Using the CR parametrization with the first level as the reference category, then the regression line for the reference category simplifies to

$$\mu_i = \beta_0 + \beta_1 X_i$$

providing a straightforward interpretation for $\beta_0$ which corresponds to the expected $Y$ for the reference group when $X = 0$. Using similar arguments we can interpret $\alpha_\ell (\ell > 2)$ as the expected difference of $Y$ between a subject of group $\ell$ and a subject in the reference category having the same $X$ since

$$E(Y|X, A = \ell) - E(Y|X, A = 1) = \beta_0 + \beta_1 X + \alpha_\ell - \beta_0 - \beta_1 X = \alpha_\ell.$$ 

Similarly, $\beta_1$ is the expected difference in $Y$ when comparing two subjects differing by one unit in $X$ and belonging in the same group since

$$E(Y|X = x+1, A = \ell) - E(Y|X = x, A = \ell) = \beta_0 + \beta_1 (x+1) + \alpha_\ell - \beta_0 - \beta_1 x - \alpha_\ell = \beta_1.$$ 

Interpretation for STZ parametrization is similar as in CR but our comparisons are can be based on the property

$$\frac{1}{L_A} \sum_{\ell=1}^{L_A} E(Y|X, A = \ell) = \beta_0 + \beta_1 X.$$ 

The above equation indicates that in STZ parametrization we assume as a reference category one which represents the average behavior across all factor levels. Following the above, $\beta_0$ under STZ parametrization provides an estimate of the expected value of $Y$ when $X = 0$ averaged across all levels of $A$ while $\alpha_\ell$ provide the expected deviation of $Y$ for a subject belonging in $\ell$ group with $X = x$ from the corresponding expected value of $Y$ averaged across all levels of $A$ having also $X = x$. The STZ parametrization is equivalent to using covariates centered around zero (i.e. using $Z_i = X_i - \bar{X}$) where the constant parameter $\beta_0$ represents the expected $Y$ for a “typical” individual of the observed data; see page 160 for more details.

The main property of this model is that the effect of $X$ is the same in each group (equal to $\beta_1$) and the effect of the grouping variable is the same when subjects of the same $X$ are compared. This is not the case in the ‘different slope’ model in which the effect of each covariate on $Y$ depends on the values of the other covariate. Graphical representation of this model is provided in Figure 5.8..
Eliminating the constant parameter $\beta_0$ from the model usually does not have a practical value. It is equivalent to assuming that expected value of the reference category (CR) or the ‘average’ across all categories (STZ) for $X = 0$ is equal to zero. Finally, centering the quantitative variable around zero simplifies the interpretation of $\beta_0$ in some cases. The constant parameter will now represent the expected $Y$ for a ‘typical’ subject belonging in the reference category (in CR parametrization).

**Figure 5.8.** Graphical representation of the parallel line model using corner constrained parameters.

It is relatively straightforward to write the likelihood of the above model in WinBUGS. We can either use the direct approach, having the data in three vectors for $y$, $x$ and $a$ or use the design matrix of the model as described in the previous sections. In the first approach the likelihood in WinBUGS can be written as

```plaintext
for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- beta0 + beta1 * x[i] + alpha[ a[i] ]
}
```

The constraints can be easily incorporated in the model using the same commands as in ANOVA models i.e.

```
alpha[1] <- 0.0
```

for CR constraints and

```
```

for STZ constraints. We can avoid imposing constraints by expressing the linear predictor as in (5.21) and hence write directly
mu[i] <- beta.star[ a[i] ] + beta1 * x[i]

Finally, when using dummy variables the linear predictor is written in the following form

$$\mu_i = \beta_0 + \beta_1 X_i + \alpha_2 D_{i2} + \ldots + \alpha_{LA} D_{iLA}$$

which can be easily coded in WinBUGS using the following expression to define the linear predictor


or, alternatively, using the inprod command

mu[i]<- beta0 + beta1*x[i] + inprod( alpha[2:LA], D[i,2:LA] )

In the above syntax, we additionally need to define alpha[1]. The inprod syntax can be also generalized for all variables included in the model as in the usual regression model. In such case all data must be stored in a data matrix X of dimension $n \times (LA + 1)$.

5.5.3.3 The separate lines model.

When assuming different regression lines for each level of the categorical variable the mean of $Y$ can be defined using the expression

$$\mu_i = \beta_0 + \beta_1 X_i + \alpha_{a_i} + \delta_{a_i} X_i.$$  \hspace{1cm} (5.22)

Alternatively, the model can be rewritten as

$$\mu_i = \beta_{0,a_i} + \beta_{1,a_i} X_i$$  \hspace{1cm} (5.23)

where $\beta_{0,\ell} = \beta_0 + \alpha_\ell$ and $\beta_{1,\ell} = \beta_1 + \delta_\ell$ for $\ell = 1, 2, \ldots, LA$. Thus, this model is equivalent to fitting one regression model for each group.

Adopting a single ANCOVA model for the whole dataset instead of fitting different regression lines for each group allows us to monitor and/or check for deviations from specific assumptions such as, for example, the common slopes assumption. Parameters $\delta_\ell$ refer to the interaction parameters between variable $X$ and factor $A$. Their magnitude reflects deviations from the parallel lines model and encapsulates changes of the effect of the quantitative covariate $X$ on $Y$ across the groups of factor $A$. STZ or CR parameterization may also be imposed to get an identifiable model.

The parameters involved in the linear predictor are given by

$${\boldsymbol{\beta}} = (\beta_0, \beta_1, \alpha_2, \ldots, \alpha_{LA}, \delta_2, \ldots, \delta_{LA})$$

Under the CR parametrization, constraints $\alpha_1 = \delta_1 = 0$ are imposed resulting in

$$\mu_i = \beta_0 + \beta_1 X_i$$

for the first-reference category. Hence parameters $\beta_0$ and $\beta_1$ will have the usual regression interpretation referring to subjects of the reference category.
The main effects $\alpha_\ell$ denote the difference in the expected values of $Y$ between subjects belonging in the $\ell$-th and the reference category when $X = 0$ since
\[
E(Y|X = 0, A = \ell) - E(Y|X = 0, A = 1) = \beta_0 + \alpha_\ell - \beta_0 = \alpha_\ell.
\]
Finally, $\beta_1 + \delta_\ell$ is the expected difference of $Y$ when comparing two subjects differing by one unit in $X$ and belonging in the $\ell$-th group since
\[
E(Y|X = x, A = \ell) - E(Y|X = 0, A = \ell) = \beta_0 + \beta_1(x + 1) + \alpha_\ell + \delta_\ell(x + 1)
- [\beta_0 + \beta_1 x + \alpha_\ell + \delta_\ell x] = \beta_0 + \delta_\ell.
\]
Therefore $\delta_\ell$ provides the difference between the effects of one unit increase of $X$ on $Y$ in the $\ell$-th group and the reference category.

Similarly, we can interpret model parameters in the STZ parametrization comparing all parameters to the ‘mean’ regression line (averaged over all groups) given by
\[
\frac{1}{L_A} \sum_{\ell=1}^{L_A} E(Y|X, A = \ell) = \beta_0 + \beta_1 X.
\]
Parameters $\alpha_\ell$ and $\delta_\ell$ will now represent differences for the intercept (i.e. for $X = 0$) and for the slope of each group from this ‘mean’ regression line.

The general ANCOVA model can be fitted in WinBUGS using the following syntax
\[
\text{mu}[i] \leftarrow \beta_0 + \beta_1 x[i] + \alpha[ a[i] ] + \delta[ a[i] ] * x[i]
\]
for the linear predictor expression (5.22). The intercept and the slope of each regression line can be calculated using the syntax

Figure 5.9. Graphical representation of the separate regression lines model using corner constrained parameters.
for (l in 1:LA){
  beta0.star[k] <- beta0 + alpha[l]
  beta1.star[k] <- beta0 + delta[l]
}

Constraints for \( \alpha_1 \) and \( \delta_1 \) must be imposed as usually. Alternatively we may directly use equation (5.23) by writing

\[
\mu[i] \leftarrow \text{beta0.star[ a[i] ] + beta1.star[ a[i] ] } * x[i]
\]

The latter has the advantage that no constraints are needed since the model is directly identifiable. The original parameters of (5.22) can be extracted by adding the following commands in our model code

\[
\text{beta0 <- beta0.star[1]}
\text{beta1 <- beta1.star[1]}
\text{for (l in 2:LA){}
  alpha[l] <- beta0.star[l] - beta0
  delta[l] <- beta1.star[l] - beta1
}
\]

for CR parametrization. For STZ parametrization we need to define \( \beta_0 \) and \( \beta_1 \) as the mean of all intercepts and slopes respectively

\[
\text{beta0 <- mean(beta0.star[1:LA])}
\text{beta1 <- mean(beta1.star[1:LA])}
\]

followed by the loop used for the specification of the CR parametrization.

Using dummies is also straightforward since the multiplicative property (5.19) also holds for the interaction term between a quantitative and a qualitative variable. Hence each interaction parameter \( \delta_\ell \) will be the coefficient of a variable defined as

\[
D_{AX}^{\ell} = D_{AX}^{\ell}X_i.
\]

The model can now be written

\[
\mu_i = \beta_0 + \beta_1X_i + \sum_{\ell=2}^{LA} \alpha_\ell D_{A \ell}^A + \sum_{\ell=2}^{LA} \delta_\ell D_{AX}^{\ell}
\]

which can be coded in WinBUGS using the syntax

\[
\mu[i]<- \text{beta0 + beta1*x[i] + alpha[2]*DA[i,2] +...+ alpha[LA]*DA[i,LA]} + \delta[2]*DA[i,2]*x[i] +...+ \delta[LA]*DA[i,LA]*x[i]
\]

The inprod command can be used to simplify the linear predictor

\[
\mu[i]<- \text{beta0 + beta1*x[i] + inprod(alpha[2:LA]*DA[i,2:LA]) + inprod(delta[2:LA]*DAX[i,2:LA])}
\]

where DAX is a matrix with each column representing the interaction term between factor \( A \) and variable \( X \) and can be calculated using the following loop

\[
\text{for (l in 1:LA){ DAX[i,l] <- DA[i,l] * x[i] }}
\]
placed nested within the likelihood loop (i.e. nested within a loop with \(i\) taking values from 1 to \(n\)). Use of the design matrix, in combination with the `\texttt{inprod}` command, further simplifies the model code as in usual regression models.

Finally removing the main effects \(\alpha_{\ell}\) will result to the common intercept model which assumes equal expected values of \(Y\) for all groups when \(X = 0\). If this model is combined with centering \(X\) round zero, then it assumes that people with average \(X\) are expected to have equal \(Y\) in all groups. Common intercept models are not frequently used in practice and they are usually adopted only if a scientific scenario supports them; see for an example in section 5.5.4.2.

### 5.5.4 A Bioassay example

Bioassays are experiments performed in the pre-clinical stage of a drug experiment. They ensure drug safety and determine appropriate drug dosage. The main objective is to estimate the unknown concentration of a new drug in comparison to standard preparation (drug) of known concentration. This unknown concentration is called potency of a drug and is usually estimated by the ratio of the mean concentration of a standard treatment over the mean of the (new) test preparation i.e.

\[
POTENCY = \frac{\text{mean concentration of std. treatment}}{\text{mean concentration of test treatment}}
\]

The mean corresponds to a response variable that may measure volume, weight or dosage of each treatment required to achieve the same results. If the relative potency is greater than one then a lower dose of the new drug produces the same results as a higher dosage of the standard drug. Therefore, the new drug is more potent than the standard treatment.

The response is usually quantitative or binary. In the following example we deal with a indirect bioassay in which we replicate an experiment in prespecified levels of dosage for the two treatments. In such assays, one or more drugs at different dosage levels are administered to experimental units such as cell-cultures, tissues, organs or living animals. The response is quantitative, hence ANCOVA is used to analyze the data and estimate the relative potency. The aim is, using statistical models, to estimate the relative potency of the new preparation. Two different approaches are popular: the parallel line analysis and the slope ratio analysis. Details concerning the parallel and slope ratio analysis can be found in Chen (2007) and references therein. Bayesian approaches to the parallel line and slope ratio analysis have been investigated by Darby (1980) and Mendoza (1990) respectively.

#### EXAMPLE 5.4

**Factor 8 example.** Factor 8 (F8) is a blood clotting agent. Deficiency of F8 in human system leads to haemophilia. The (fictitious) data of an experiment are given in Table 5.23, in which we wish to compare a new/test preparation with the standard preparation of F8 having potency equal to 1.2 international units (i.u.). As a response we use the time until clotting occurs in seconds. Three dilutions (doses) were used (1:40, 1:20, 1:10) with four measurements per doses.

#### 5.5.4.1 Parallel Line Analysis.

**Model Formulation.** For a fixed dose \(d_s\) of standard drug, we assume that the same effect is achieved with dose \(d_t = \rho d_s\) of the test preparation; where \(\rho\) is the relative potency of
INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

<table>
<thead>
<tr>
<th>Dose</th>
<th>Blood clotting time (in seconds)</th>
<th>Standard</th>
<th>Test (new)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:40</td>
<td>0.025</td>
<td>68.8, 67.6, 68.1, 67.6</td>
<td>69.0, 67.9, 68.6, 68.3</td>
</tr>
<tr>
<td>1:20</td>
<td>0.050</td>
<td>61.4, 59.8, 62.3, 60.6</td>
<td>60.9, 60.3, 61.6, 61.8</td>
</tr>
<tr>
<td>1:10</td>
<td>0.100</td>
<td>53.5, 51.9, 53.6, 52.2</td>
<td>53.8, 54.9, 54.1, 54.2</td>
</tr>
</tbody>
</table>

Table 5.23. Data for Bioassay example (example 5.4).

the two drugs. We adopt the model

\[
E(Y) = \mu = \begin{cases} 
\beta_0 + \beta_1 \log(\text{dose}) & \text{for the standard treatment} \\
\beta_0 + \beta_1 \log(\rho \times \text{dose}) & \text{for the test treatment}
\end{cases}
\]

Figure 5.10. Graphical representation of the parallel line analysis model.

The model for the test treatment can be rewritten as

\[
\mu_T = \beta_0 + \beta_1 \log(\rho \times \text{dose}) = \beta_0 + \beta_1 \log(\rho) + \beta_1 \log(\text{dose}) = \beta'_0 + \beta_1 \log(\text{dose}),
\]

where \( \beta'_0 = \beta_0 + \beta_1 \log(\rho) \). The two models have the same slope but different intercept. The common slope is given by \( \beta_1 \) while the intercept of the standard treatment by \( \beta_0 \) and for the test treatment by \( \beta'_0 \). From the above, the relative potency can be calculated by

\[
\beta'_0 = \beta_0 + \beta_1 \log(\rho) \Leftrightarrow \rho = \exp\left(\frac{\beta'_0 - \beta_0}{\beta_1}\right). \tag{5.24}
\]

Using MCMC, it is straightforward to calculate the posterior distribution of \( \rho \) either directly by including it in the model by simply calculating the above quantity from the values of a parallel line ANCOVA model.
In Table 5.24, we provide the WinBUGS code for fitting the parallel lines model. In this model code different intercepts have fitted directly to avoid contraints. Alternatively, the code using the design matrix approach for the CR parametrization is given in Table 5.25. Commands for fitting the model using STZ parametrization are also provided as code comments in the same Table. A normal prior distribution with mean zero and variance equal to 1000 was used for the parameters of the linear predictor and gamma with mean one and variance 1000 for the model’s precision.

```winbugs
model{
  # model’s likelihood
  for (i in 1:n){
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- beta0[drug[i]] + beta1*log(dose[i])
  }
  # relative potency
  rho <- exp((beta0[2]-beta0[1])/beta1)
  # potency estimate
  potency <- rho * 1.2
  # prior distributions
  beta0[1] ~ dnorm( 0.0, 0.001) # constant for standard treatment
  beta0[2] ~ dnorm( 0.0, 0.001) # constant for test treatment
  beta1 ~ dnorm( 0.0, 0.001) # slope
  tau ~ dgamma( 0.001, 0.001) # precision of regression model
  s <- 1/sqrt(tau) # standard error of regression
  # test rho>1 (test more potent)
  more.potent21 <- step(rho-1)
  #
  # intercept.difference <- beta0[2]-beta0[1]
  }
}

INITS
list( beta0=c(0,0), beta1=0, tau=1 )

DATA (LIST)
list( n=24,
  y=c(68.8, 67.6, 68.1, 67.6, 69.0, 67.9, 68.6, 68.3,
    61.4, 59.8, 62.3, 60.6, 60.9, 60.3, 61.6, 61.8,
    53.5, 51.9, 53.6, 52.2, 53.8, 54.9, 54.1, 54.2),
  dose=c(0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025,
    0.050, 0.050, 0.050, 0.050, 0.050, 0.050, 0.050, 0.050,
    0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100, 0.100),
  drug=c(1,1,1,1,2,2,2,2, 1,1,1,1,2,2,2,2, 1,1,1,1,2,2,2,2) )
```

Table 5.24. WinBUGS code for the parallel line model for example 5.4 (Bioassay Data).

Results. Posterior summaries of the parallel lines model have been calculated (see Table 5.26.) after running the MCMC algorithm for 2000 iterations and discarding additional 1000
INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

model{
  for (i in 1:n){

    # creating the design matrix
    X[i,1]<-1.0 # beta1=constant term
    X[i,2]<-log(dose[i]) # beta2=log dose
    X[i,3]<- equals( drug[i], 2 ) # beta3=CR dummy for test treatment
    #X[i,3]<- equals( drug[i], 2 )-equals( drug[i], 1 ) # beta3=STZ dummy

    model likelihood
    y[i] ~ dnorm( mu[i], tau )
    mu[i] <- inprod( beta[], X[i,])
  }

  #
  #rho <- exp( 2*beta[3]/beta[2] ) # relative potency in STZ
  # potency estimate
  potency <- rho * 1.2
  #
  # prior distributions
  for (j in 1:3){ beta[j]~dnorm( 0.0, 0.001) }
  tau ~ dgamma( 0.001, 0.001) # precision of regression model
  s <- 1/sqrt(tau) # standard error of regression
}

INITS
list( beta=c(0,0,0), tau=1 )

Table 5.25. WinBUGS code for the parallel line model for example 5.4 (Bioassay Data) using the design matrix approach; data are specified as in Table 5.24.; (substitute lines 7 and 14 by lines 8 and 15 respectively to switch from CR to STZ parametrization).
iterations as a burn-in period. The following point estimate of the model can be provided using the posterior means:

\[ y_i \sim N(\mu_i, 0.77^2), \quad \mu_i = \begin{cases} 
28.79 - 10.62 \log(\text{dose}) & \text{for the standard treatment} \\
29.46 - 10.62 \log(\text{dose}) & \text{for the test treatment}
\end{cases} \]

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0[1]</td>
<td>28.79</td>
<td>0.865</td>
<td>0.01804</td>
<td>27.09</td>
<td>28.8</td>
<td>30.45</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>beta0[2]</td>
<td>29.46</td>
<td>0.8701</td>
<td>0.01875</td>
<td>27.75</td>
<td>29.46</td>
<td>31.2</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>beta1</td>
<td>-10.62</td>
<td>0.2808</td>
<td>0.006079</td>
<td>-11.18</td>
<td>-10.62</td>
<td>-10.05</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>potency</td>
<td>1.127</td>
<td>0.03341</td>
<td>7.865E-4</td>
<td>1.062</td>
<td>1.127</td>
<td>1.195</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>rho</td>
<td>0.939</td>
<td>0.02785</td>
<td>6.554E-4</td>
<td>0.8846</td>
<td>0.9391</td>
<td>0.9961</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>s</td>
<td>0.7731</td>
<td>0.1318</td>
<td>0.003941</td>
<td>0.5753</td>
<td>0.7532</td>
<td>1.085</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>tau</td>
<td>1.81</td>
<td>0.5676</td>
<td>0.01645</td>
<td>0.8504</td>
<td>1.763</td>
<td>3.027</td>
<td>1001</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 5.26. Posterior summaries for the parameters of the parallel line model in example 5.4 (Bioassay Data).

Looking at the estimates of the model we conclude to the following

- According to the intercepts \( \beta_0 \) and \( \beta_0' \), the clotting time when the dose is equal to one \((\log\text{-dose}=0)\) is a-posteriori expected to be equal to 28.8 and 29.5 seconds respectively. This interpretation does not have any practical meaning since this values is far away from the range of dose values used in this experiment. To get sensible estimates we may transform the log-dose such that the zero value corresponds to a realistic dosage level (for example the lower level of dosage); see Table 5.27. for new rescaled results.

- We observe negative association between the drug dosage and the clotting time. Interpretation is not exactly the same as in usual regression models in terms of dosage increase since the log-dose is used as an explanatory variable. One unit increase of the log-dose corresponds to multiplying the original dosage by \( e \approx 2.7 \) since

\[
E(Y|X = \log(\text{dose}) + 1, A = k) - E(Y|X = \log(\text{dose}), A = k) = \beta_1
\]

and

\[
\log(\text{new dose}) = \log(\text{dose}) + 1 = \log(\text{dose}) + \log e = \log(e \times \text{dose}) \Rightarrow \text{new dose} = e \times \text{dose}.
\]

Hence we can now interpret \( \beta_1 \) as the expected change in clotting time when the dosage increases by 1.7 times (170%) the original one. In our data, when the dosage increases by 170%, the clotting time is a-posteriori expected to decrease by 10.6 seconds.

To get more interpretable parameters, we propose to use as explanatory variable the \( \log_2(40 \times \text{dose}) \) instead of the \( \log(\text{dose}) \). By this way the rescaled explanatory variable will take values equal to 0, 1, 2. Now the intercept is directly linked with the expected clotting time of the lower dosage used in the experiment while the new slope coefficient is associated with the expected decrease of clotting time when the dosage becomes double. Results of this rescaled model are provided in Table 5.27.
In order to calculate the proposed transformation in WinBUGS we need to express the new variable in terms of \( \log(x) \). We write

\[
x = \log_2(40 \times \text{dose}) = \frac{\log(40 \times \text{dose})}{\log(x)}
\]

using the command

\[
x[i] <- \frac{\log(40 \times \text{dose}[i])}{\log(x[i])}
\]

and then use \( x[i] \) in the linear predictor instead of \( \log(\text{dose}[i]) \). A slight change is also needed for the calculation of the relative potency since the new model is now given by

\[
\mu_i = \beta_{0a_i} + \beta_1 \frac{\log(40 \times \text{dose}_i)}{\log(2)}
\]

\[
= \beta_{0a_i} + \beta_1 \frac{\log(40)}{\log(2)} - \beta_1 \frac{\log(\text{dose}_i)}{\log(2)}.
\]

The original parameters \( \beta_{0k} \) and \( \beta_1 \) are associated with the parameters of the new rescaled model by the equations

\[
\beta_{0k} = \beta_{0k} + \beta_1 \frac{\log(40)}{\log(2)} \quad \text{and} \quad \beta_1 = \frac{\beta_1}{\log(2)}
\]

resulting to relative potency

\[
\rho = \exp\left(\frac{\beta_{02} - \beta_{01}}{\beta_1}\right) = \exp\left(\log(2) \frac{\beta_{02} - \beta_{01}}{\beta_1}\right) = 2 \left(\frac{\beta_{02} - \beta_{01}}{\beta_1}\right).
\]

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0[1]</td>
<td>67.96</td>
<td>0.3024</td>
<td>0.006875</td>
<td>67.37</td>
<td>67.96</td>
<td>68.55</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>beta0[2]</td>
<td>68.63</td>
<td>0.2973</td>
<td>0.006413</td>
<td>68.05</td>
<td>68.63</td>
<td>69.23</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>beta1</td>
<td>-7.353</td>
<td>0.1971</td>
<td>0.004279</td>
<td>-7.736</td>
<td>-7.354</td>
<td>-6.973</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>potency</td>
<td>1.127</td>
<td>0.03467</td>
<td>8.547E-4</td>
<td>1.058</td>
<td>1.126</td>
<td>1.198</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>rho</td>
<td>0.939</td>
<td>0.02889</td>
<td>7.123E-4</td>
<td>0.8816</td>
<td>0.9386</td>
<td>0.9981</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>s</td>
<td>0.7731</td>
<td>0.1318</td>
<td>0.003943</td>
<td>0.5754</td>
<td>0.7532</td>
<td>1.086</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>inter.dif</td>
<td>0.654</td>
<td>0.307</td>
<td>0.007264</td>
<td>0.0543</td>
<td>0.6648</td>
<td>1.239</td>
<td>1001</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 5.27. Posterior summaries for the parameters of the parallel line model after rescaling the dose (\( \log_2(40 \times \text{dose}) \)) in example 5.4 (Bioassay Data).

The interpretation of the model parameters is more straightforward since

- The clotting time for the lower dosage is a-posteriori expected to be equal to 68 and 68.6 seconds for the standard and the test treatment respectively. The difference between the two drugs is about 0.65 seconds with about 98% of the posterior values to be positive.
- When we double the dosage we a-posteriori expect decrease of the clotting time by 7.4 seconds.
- From the posterior densities for the two intercepts (see Figure 5.11.) we observe that they lie in the same range of values deducing equal potency between the two drugs.
This comparison might be misleading since the two variables highly are correlated (see Figure 5.12.). To get more reliable picture we consider the posterior density of their difference ($\beta'_0 - \beta_0$). Indeed from the posterior density of the difference (see Figure 5.13.) we observe that the values of zero (equal intercepts) lies at the right tail area of the posterior distribution of the difference being equal to the 1.85% percentile of the posterior distribution. This indicates that the two treatments differ in terms of potency.

![Figure 5.11. Posterior Box-plots for the intercepts under the original and the rescaled parallel lines models in example 5.4 (Bioassay Data).](image)

![Figure 5.12. Posterior scatterplots for the intercepts under the original and the rescaled parallel lines models in example 5.4 (Bioassay Data).](image)

Similar are the conclusions if we monitor directly the relative potency $\rho$ since it is a simple function of the above difference. We observe that the test treatment is a-posteriori expected to be about 6% less potent than the standard treatment ranging from 0.88 to 0.998 with probability 95%. The actual potency of the test treatment was a-posteriori expected to be equal to 1.13 i.u. ranging from 1.058 to 1.198 i.u. with probability equal to 95%.

**Checking the parallel lines assumption.** We may check the parallelism assumption by simply fitting the separate lines model and comparing the posterior distribution of the
INCORPORATING CATEGORICAL VARIABLES IN NORMAL MODELS.

(a) original model [using $\log(dose)$]  
(b) rescaled model [using $\log_2(40 \times dose)$]

**Figure 5.13.** Posterior densities for the intercepts difference under the original and the rescaled parallel lines models in example 5.4 (Bioassay Data).

**Figure 5.14.** Posterior density of the relative potency the new compared to the standard treatment using the parallel line model in example 5.4 (Bioassay Data).
differences between slopes. In order to fit this model in WinBUGS we simply change the linear predictor to

\[ \mu_i = \beta_0[\text{drug}[i]] + \beta_1[\text{drug}[i]] \times \log(\text{dose}[i]) \]

and define normal prior distributions for \( \beta_0[1], \beta_0[2], \beta_1[1] \) and \( \beta_1[2] \). We also define the slope difference using the command

\[ \text{slope.difference} <- \beta_1[2] - \beta_1[1] \]

is order to monitor its posterior distribution and

\[ p_2 <- 1 - \text{step}(\text{slope.difference}) \]

which calculates the posterior probability that the slope difference is lower than zero value (parallel lines assumption). Very high or low values provide an indication that these slopes are different. Results using 2000 iterations (using the rescaled log-dose) are given in Table 5.28.; see also Figure 5.15. for the posterior density plot. From these results we observe that the a-posteriori expected slope difference is equal to 0.51 ranging from -0.22 to 1.28 with probability 95%. We also estimate \( f(\text{slope difference} < 0) = 0.084 \) indicating that, although the zero value lies at the left tail area of the posterior distribution, it does not provide strong evidence against the parallel lines assumption.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2</td>
<td>0.084</td>
<td>0.2774</td>
<td>0.005963</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1001</td>
<td>2000</td>
</tr>
<tr>
<td>slope.difference</td>
<td>0.5096</td>
<td>0.3783</td>
<td>0.008266</td>
<td>-0.2171</td>
<td>0.4961</td>
<td>1.277</td>
<td>1001</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 5.28. Posterior summaries for the parameters of the separate lines model in example 5.4 (Bioassay Data) using \( \log(40 \times \text{dose})/\log(2) \) as explanatory variable.

Figure 5.15. Posterior density of the slope difference in the separate lines model for example 5.4 (Bioassay Data).

### 5.5.4.2 Slope Ratio Analysis
Incorporating Categorical Variables in Normal Models.

Model formulation. In this approach we assume again that for any fixed dose \(d_s\) of the standard preparation we have the same effect with the dose \(d_t = \rho d_s\) of the test preparation, but now we adopt the model

\[
E(Y) = \mu = \begin{cases} 
\beta_0 + \beta_1 \text{dose} & \text{for the standard treatment} \\
\beta_0 + \beta_1 \rho \text{dose} & \text{for the test treatment} 
\end{cases}
\]

The model for the test treatment can be rewritten as

\[
\mu_T = \beta_0 + \beta_1' \text{dose}
\]

where \(\beta_1' = \beta_1 \rho\). The two models have now the same intercept but different slope. The common intercept is given by \(\beta_0\) while the slope of the standard treatment by \(\beta\) and for the test treatment by \(\beta_1'\). From the above, we can calculate the relative potency by

\[
\rho = \frac{\beta_1'}{\beta_1}.
\]

(5.25)

It is straightforward to calculate the relative potency when the common intercept model is fitted using interaction terms under CR and STZ parametrization. The model can be generally expressed using the linear predictor

\[
\mu_i = \beta_0 + \beta_1 \text{dose}_i + \delta \text{dose}_i
\]

with \(\delta_1 = 0\) and \(\delta_1 = -\delta_2\) for CR and STZ parameterizations respectively. The slopes \(\beta_0\) and \(\beta_0'\) are equal to \(\beta_1\) and \(\beta_1 + \delta_2\) in CR resulting to

\[
\rho = \frac{\beta_1 + \delta_2}{\beta_1}
\]

while for STZ \(\beta_0 = \beta_1 - \delta_2\) and \(\beta_0' = \beta_1 + \delta_2\) with relative potency given by

\[
\rho = \frac{\beta_1 + \delta_2}{\beta_1 - \delta_2}
\]

Before we proceed, we rescale the dose in order to attain a model with parameters of simple interpretation. For this reason, in this analysis, we use the transformation

\[
x = \frac{\text{dose}}{\text{min(dose)}}.
\]

(5.26)

By this way, the two regression slopes provide the expected decrease of the clotting time when the dosage is increased by a quantity equal to the minimum dose used in the experiment. Note that, in this analysis we must not change the zero point of the new transformed variable since such a transformation has a direct effect on the estimated model. For example transformation

\[
x = \frac{\text{dose} - \text{min(dose)}}{\text{min(dose)}}
\]

results in \(x = 0\) for dosage equal to the minimum dose of the experiment. Under this transformation, the common intercept \(\beta_0\) of the model corresponds to the expected clotting time for the lower dose. Therefore this model assumes equal effect of the lower dose for both treatments instead of the more realistic assumption of equal effect for the zero dosage of both treatments imposed by the original slope ratio analysis model.
Figure 5.16. Graphical Representation of the Slope Ratio Analysis Model.

Under transformation (5.26), the relative potency is still given by the ratio of the two slopes. The minimum value of a vector is calculated in WinBUGS using the command `ranked(dose[], 1)`. Hence the command

```plaintext
x[i] <- dose[i]/ranked(dose[], 1)
```

calculates the new rescaled dose used in the linear predictor of our model.

WinBUGS code for the common intercept model is provided in Table 5.29, using the rescaled dosage (5.26). Different slopes have been used to avoid constraints. Alternatively, our model can be constructed using the design matrix approach as in the parallel lines model. The corresponding code for the CR parametrization is provided in Table 5.30. Commands for fitting the model using STZ parametrization are also provided in the same Table as comments within the model code. Finally, syntax for specifying improper flat priors \( f(\beta_j) \propto 1 \) are also provided as model code comments. These improper priors were used to check the sensitivity of the posterior distribution. Minor differences were observed between the posteriors resulted by the improper and the \( N(0, 10^3) \) prior distributions.

**Results.** A normal prior distribution with mean zero and variance equal to 1000 was used for the parameters of the linear predictor and gamma with mean one and variance 1000 for the model’s precision. Results using 2000 iterations and discarding the initial 1000 iterations are provided in Table 5.31. When using the original dose, estimated posterior means and variances are found large for specific parameters. The above normal prior distributions were proved to be informative since results were sensitive to different values of the prior variance. In this case, the flat improper distribution (using the command `dflat()`) was used.

A point estimate of the model based the posterior means is given by

\[
y_i \sim N(\mu_i, 0.77^2), \quad \mu_i = \begin{cases} 
71.98 - 4.88x_i & \text{for the standard treatment} \\
71.98 - 4.58x_i & \text{for the test treatment}
\end{cases}
\]
model{
    for (i in 1:n){
        # calculate the rescaled dose
        x[i] <- dose[i]/ranked(dose[],1)
        # model's likelihood
        y[i] ~ dnorm( mu[i], tau )
        mu[i] <- beta0 + beta1[drug[i]]*x[i]
    }
    
    # relative potency
    rho <- beta1[2]/beta1[1]
    # potency estimate
    potency <- rho * 1.2
    
    # prior distributions
    # normal priors
    beta0 ~ dnorm( 0.0, 0.001) # constant for standard treatment
    beta1[1] ~ dnorm( 0.0, 0.001) # constant for test treatment
    beta1[2] ~ dnorm( 0.0, 0.001) # slope
    ## flat improper priors
    #beta0 ~ dflat() # constant for standard treatment
    #beta1[1] ~ dflat() # constant for test treatment
    #beta1[2] ~ dflat() # slope
    tau ~ dgamma( 0.001, 0.001) # precision of regression model
    s <- 1/sqrt(tau) # standard error of regression
    
    # test rho>1 (test more potent)
    more.potent21 <- step(rho-1)
}
INITS
list( beta1=c(0,0), beta0=0, tau=1 )

Table 5.29. WinBUGS code for the parallel line model for example 5.4 (Bioassay Data).
model{
    for (i in 1:n){
        # creating the design matrix
        X[i,1]<-1.0 # beta1=constant term
        # beta2=rescaled dose for standard treatment
        X[i,2]<-dose[i]/ranked(dose,1)
        # beta3=CR dummy for interaction (slope difference)
        X[i,3]<- equals( drug[i], 2 )*X[i,2]
        # model likelihood
        y[i] ~ dnorm( mu[i], tau )
        mu[i] <- inprod( beta[], X[i,])
    }
    # potency estimate
    potency <- rho * 1.2
    # prior distributions
    for (j in 1:3){ beta[j]~dnorm( 0.0, 0.0001) } # normal priors
    for (j in 1:3){ beta[j]~dflat() } # flat improper priors
    tau ~ dgamma( 0.001, 0.001) # precision of regression model
    s <- 1/sqrt(tau) # standard error of regression
}

INITS
list( beta=c(0,0,0), tau=1 )

Table 5.30. WinBUGS code for the common intercept model for example 5.4 (Bioassay Data) using the design matrix approach and rescaled dose; data are specified as in Table 5.24.; (substitute lines 8-9 and 17 by lines 10-11 and 18 respectively to switch from CR to STZ parametrization).

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error 2.5%</th>
<th>median 97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>71.98</td>
<td>0.6029</td>
<td>0.01443</td>
<td>71.99</td>
<td>71.99</td>
<td>1001</td>
</tr>
<tr>
<td>beta1[2]</td>
<td>-4.58</td>
<td>0.2508</td>
<td>-5.078</td>
<td>-4.58</td>
<td>-4.065</td>
<td>1001</td>
</tr>
<tr>
<td>more.potent21</td>
<td>0.0805</td>
<td>0.2721</td>
<td>0.005348</td>
<td>0.0</td>
<td>0.0</td>
<td>1001</td>
</tr>
<tr>
<td>potency</td>
<td>1.126</td>
<td>0.05168</td>
<td>0.001109</td>
<td>1.027</td>
<td>1.125</td>
<td>1001</td>
</tr>
<tr>
<td>rho</td>
<td>0.9382</td>
<td>0.05168</td>
<td>0.001109</td>
<td>0.8555</td>
<td>0.9378</td>
<td>1001</td>
</tr>
<tr>
<td>s</td>
<td>1.362</td>
<td>0.2325</td>
<td>0.006969</td>
<td>1.014</td>
<td>1.33</td>
<td>1001</td>
</tr>
<tr>
<td>tau</td>
<td>0.5829</td>
<td>0.1828</td>
<td>0.005289</td>
<td>0.2727</td>
<td>0.5657</td>
<td>1001</td>
</tr>
</tbody>
</table>

Table 5.31. Posterior summaries for the parameters of the common slope model in example 5.4 (Bioassay Data); rescaled dose (5.26) is used as explanatory variable.
where $x_i$ is given by (5.26). From the above we can infer that

- The clotting time is a-posteriori expected to be equal to 72 seconds when no dilution is used.

- The clotting time is a-posteriori expected to be reduced by 4.9 seconds when increasing the dose of the standard treatment by its minimum quantity (1:40) used in the experiment.

- The clotting time is a-posteriori expected to be reduced by 4.6 seconds when increasing the dose of the test treatment by its minimum quantity (1:40) used in the experiment.

- The new (test) treatment is a-posteriori expected to be $6.2\%$ less potent than the standard treatment with values ranging from 0.86 to 1.03 with 95% probability. Although the posterior mean of the relative potency is similar to the corresponding one in the parallel analysis, the 95% posterior interval is wider in the current analysis including the value of one which correspond to the assumption of equal treatment potency. The posterior probability that the second treatment is more potent than the standard one was found equal to 0.08 which is considerably higher than the corresponding probability in the parallel lines analysis.

- The new (test) treatment is a-posteriori expected to have potency equal to 1.13 international units with values ranging from 1.03 to 1.23 i.u with 95% posterior probability.

**Checking the assumption of common intercepts.** In the above model we need to check for the equal intercepts assumption. The separate lines model can be used to check this assumption. From the posterior summaries of Table 5.32, and the posterior density plot of the intercept difference in Figure 5.18, we observe that the zero value (which corresponds to the parallel lines assumption) is close to the center of the posterior distribution since the the posterior mean of the intercept difference is equal to $-0.21$ and the 95% ranges from
−2.56 to 2.18. Moreover, the zero value corresponds to the 42% percentile of the posterior distribution indicating that this assumption is a-posteriori sensible.

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept.difference</td>
<td>−0.2147</td>
<td>1.199</td>
<td>0.02616</td>
<td>−2.564</td>
<td>−0.2198</td>
<td>2.184</td>
</tr>
<tr>
<td>p2</td>
<td>0.424</td>
<td>0.4942</td>
<td>0.01217</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 5.32. Posterior summaries for the intercept difference of the separate lines model in example 5.4 (Bioassay Data); rescaled dose (5.26) is used as explanatory variable.

Figure 5.18. Posterior density of the intercept difference in the separate lines model for example 5.4 (Bioassay Data).

5.5.4.3 Discussion and comparison of the two approaches.

Two different approaches based on ANCOVA models have been implemented in order to estimate the relative potency of a new drug using a simple Bioassay example. From the above, both adopted models indicate that the new drug is about 6% less potent than the standard. The first analysis indicates clear differences between the two drugs since the posterior density of the estimated relative potency is far away from the value of one which corresponds to equally potent drugs. From the second model, the difference is not so clear since the posterior distribution of the estimated potency is more dispersed indicating that the equal potency assumption might be plausible. In order to reach our final conclusion we need to determine which model describes better the under study random behavior. For this reason we may use a ‘naive’ approach base on the Bayesian versions of $R^2$ quantities described earlier in this chapter or more sophisticated techniques based on the predictive distributions to check the assumptions of each model and its goodness of fit. Also, Bayesian model comparison techniques can be used to identify which model is more appropriate in this case. All this issues will be described in chapters 8 and 9 which follow.
Before closing we provide the Bayesian measures of $R^2_B$ in Table 5.33, in order to monitor the goodness of fit of the models. This measure indicates that the models of the parallel lines analysis fit the data slightly better than the models of the slope-ratio analysis. Moreover the non-common slopes model (in parallel analysis) only slightly increases the posterior mean of this measure (posterior densities are identical). Similar is the result in the non common intercept model in the slope-ratio analysis since the posterior mean of $R^2_B$ is slightly lower that the corresponding one in the common slope model while all posterior summaries are very close.

<table>
<thead>
<tr>
<th>model</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Parallel lines</td>
<td>0.9839</td>
<td>0.005795</td>
<td>1.715E-4</td>
<td>0.9693</td>
<td>0.9852</td>
<td>0.9914</td>
</tr>
<tr>
<td>2 Non common slope</td>
<td>0.9847</td>
<td>0.005355</td>
<td>1.341E-4</td>
<td>0.9714</td>
<td>0.9858</td>
<td>0.9919</td>
</tr>
<tr>
<td>3 Common intercept</td>
<td>0.9501</td>
<td>0.01801</td>
<td>5.352E-4</td>
<td>0.9042</td>
<td>0.9538</td>
<td>0.9732</td>
</tr>
<tr>
<td>4 Non common intercept</td>
<td>0.9479</td>
<td>0.01827</td>
<td>4.56E-4</td>
<td>0.9016</td>
<td>0.9515</td>
<td>0.9726</td>
</tr>
</tbody>
</table>

Table 5.33. Posterior summaries for $R^2_B$ in example 5.4 (Bioassay Data).

5.5.4.4 Extending the simple ANCOVA model.

In practice, the number of quantitative and qualitative variables involved in a regression model is high and the applied statistical analyst must select between them. Extending the above simple ANCOVA models by incorporating interaction terms between more than two variables or factors results in highly complicated models. Nevertheless, their interpretation can be based on the above three simple ANCOVA models (parallel lines, common intercept and separate regression lines).

When dealing with multiple regression/ANCOVA models the user must bare in mind that

- The most frequently used model is the ‘main effects’ models in which we do not include any interaction. This model corresponds to the simple model of ’parallel’ lines in the above simple ANCOVA model. It is simple to interpret and easily understood by scientists from other scientific fields.
- Generally interactions must be avoided unless the data or a scientific scenario of the problem in hand supports such an action. By avoiding interactions we simplify the model and make it easily understandable to scientists not familiar to statistics.
- When we select to use interactions in multiple regression/ANCOVA models we usually restrict in two way interactions. This is mainly due to their interpretation which is simpler than in models with higher order interactions. Moreover, the size of the data usually do not allow us to extend the model by including higher order interactions.
- An interaction between categorical variables imposes different intercepts for the level combinations of the factors involved in the corresponding interaction term.
- Higher order interaction terms which involve one quantitative variable imposes different slopes (effects) of the quantitative variable for each level combination of the factors involved in this interaction term.

Usually, model selection methods are implemented to identify well fitted models. Stepwise like methods may be used within the Bayesian context but such procedures will be intensive due to the computational effort needed to fit Bayesian models. This can be simplified
in normal models since any posterior distribution of interest can be calculated analytically when using conjugate prior distributions. Alternatively, ‘automatic’ model comparison and selection methods based on MCMC schemes can be applied. These methods are briefly presented and discussed in chapter 9.

5.6 USING BINARY INDICATORS TO SPECIFY MODELS IN MULTIPLE REGRESSION

When we wish to fit various models by trying to exclude/include different variables in the linear predictor then we need to write in WinBUGS different model codes for each model. A much simpler approach can be based on incorporating a binary vector \( \gamma = (\gamma_1, \ldots, \gamma_p) \) to the linear predictor \( \eta_i \) which can be now written as

\[
\mu_i = \beta_0 + \sum_{j=1}^{p} \gamma_j \beta_j X_{ij}
\]

and defined in WinBUGS using the syntax

\[
\text{mu}[i] = \text{beta0} + \text{gamma}[1]\text{*beta}[1]\text{*x}[i,j] + \ldots + \text{gamma}[p]\text{*beta}[p]\text{*x}[i,p]
\]

where \( X_{ij} \) (and \( x[i,j] \)) is the \( j \) covariate or dummy variable. The binary indicator \( \gamma_j \) is set equal to one if we wish to include \( X_j \) variable in our model and zero otherwise. This can be specified in the WinBUGS data section without changing the model code itself. Although we may extend the above expression using a binary indicator also for the constant term, \( \beta_0 \) is usually included in the model. Moreover, in multiple regression models with large number of covariates it is convenient to use the data matrix \( X \) in combination with the \text{inprod} command to define the model. In this case the linear predictor can be expressed as

\[
\mu_i = \sum_{j=1}^{P} \gamma_j \beta_j X_{ij} = \sum_{j=1}^{P} \beta_{\gamma,j} X_{ij}
\]

where \( \beta_{\gamma,j} = \gamma_j \beta_j \) and \( P = p + 1 \) the number of parameters involved in the linear predictor. The first column of \( X \) corresponds to the constant term while the rest of them to the quantitative or dummy variables considered as covariates.

The vector \( (\beta_{\gamma,j}) \) can be specified in WinBUGS using the syntax

\[
\text{for (j in 1:P) \{ beta.g[j] <- gamma[j]*beta[j] \}}
\]

while the linear predictor will be now given using the \text{inprod} command by

\[
\text{mu}[i] = \text{inprod( beta.g[], X[i,] )}
\]

The above binary vector \( \gamma \) is a model indicator and it is used in the bibliography related to Bayesian variable selection methods (see for example in George & McCulloch 1993). It is relatively easy to extend our model to incorporate uncertainty concerning the inclusion of each covariate. We only have to use a Bernoulli prior for each \( \gamma_j \). For example \( \gamma_j \sim \text{Bernoulli}(1/2) \) can be used to express prior indifference concerning the inclusion or exclusion of \( X_j \) variable from the model. Other technical issues such as the selection of the prior distribution and the Bartlett-Lindley paradox (Lindley 1957, Bartlett 1957) further complicate the implementation of the Bayesian variable selection. This issue is discussed in more detail in chapter 9.