Question 1:

Solution 1. *Our data is the number of people drown per year* (y) *in Finland* $1980-2013(x)$ *and we build a linear model with Gaussian noise as following:*

$$y_i = \beta_0 + \beta_1(x_i - \mu_X) + \epsilon_i, i = 1, \ldots, n$$

Here $n = 34$ *and* $\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 1996.5$, *and*

$$\epsilon_i \overset{iid}{\sim} N(0, \tau) \qquad (\tau \text{ represents precesion})$$

We compare results with three different setups:

1. *assume flat prior on* β_0 *and* β_1 ($\beta_0 \sim dflat()$ *and* $\beta_1 \sim dflat()$)*;*
2. *assume weak normal prior on* β_0 *and* β_1*, i.e.,* $\beta_0 \sim N(0, .0001)$ *and* $\beta_1 \sim N(0, .0001)$ *(here* $.0001$ *represents precision)*
3. *assume a hierarchical structure, namely,* $\beta_0 \sim N(\mu_0, \tau_0)$*,* $\beta_1 \sim N(\mu_1, \tau_1)$ *and we assume* $\mu_0 \sim dflat()$*,* $\sigma_0 \sim unif(.001, 1000)$*,* $\mu_1 \sim dflat()$ *and* $\sigma_0 \sim unif(.001, 1000)$ *(we use winBUGS language here and* σ_0*,* σ_1 *represents the standard deviation, in correspondence to the precision* τ_0 *and* τ_1*).*

Our winBUGS model setup is as following:

```
#HW5 Question 1

#Model
model{
   for (i in 1:n){
   y[i] ~ dnorm(mu[i], tau)
   mu[i] <- beta0 + beta1*(x[i] - meanX)
}
#posterior prediction for year 2016
yp ~ dnorm(mup, tau)
mup <- beta0 + beta1*(2016 - meanX)

#Priors
sig ~ dunif(.001, 1000)
tau <- 1/pow(sig, 2)

#flat prior
#beta0 ~ dflat()
#beta1 ~ dflat()

#weak normal prior
#beta0 ~ dnorm(0, .0001)
#beta1 ~ dnorm(0, .0001)

#hierarchical
beta0 ~ dnorm(mu0, tau0)
beta1 ~ dnorm(mu1, tau1)
mu0 ~ dflat()
mu1 ~ dflat()
sig0 ~ dunif(.001, 1000)
sig1 ~ dunif(.001, 1000)
tau0 <- 1/pow(sig0, 2)
tau1 <- 1/pow(sig1, 2)
}
```

We burn the first 10,000 *iterations and then update model for further* 10,000 *times, then we get the following results: For flat prior, our estimate and DIC is:*

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	220.9	5.105	0.05429	210.8	221.0	231.0	10001	10000
beta1	-3.284	0.522	0.005419	-4.311	-3.287	-2.251	10001	10000
yp	156.7	32.27	0.3032	93.32	156.7	219.6	10001	10000

DIC

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

	Dbar	Dhat	pD	DIC
y	325.399	322.493	2.906	328.306
total	325.399	322.493	2.906	328.306

For weak normal prior, our estimate and DIC is:

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	220.4	5.14	0.05053	210.2	220.4	230.5	10001	10000
beta1	-3.282	0.522	0.005359	-4.305	-3.29	-2.253	10001	10000
yp	156.4	32.24	0.3491	93.22	156.5	220.6	10001	10000

DIC

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

	Dbar	Dhat	pD	DIC
y	325.442	322.535	2.907	328.349
total	325.442	322.535	2.907	328.349

For hierarchical model, our estimate and DIC is:

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	221.1	5.108	0.04956	211.0	221.0	231.0	10001	10000
beta1	-3.285	0.5259	0.00522	-4.321	-3.279	-2.253	10001	10000
yp	157.0	32.34	0.3072	92.81	156.8	221.2	10001	10000

DIC

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

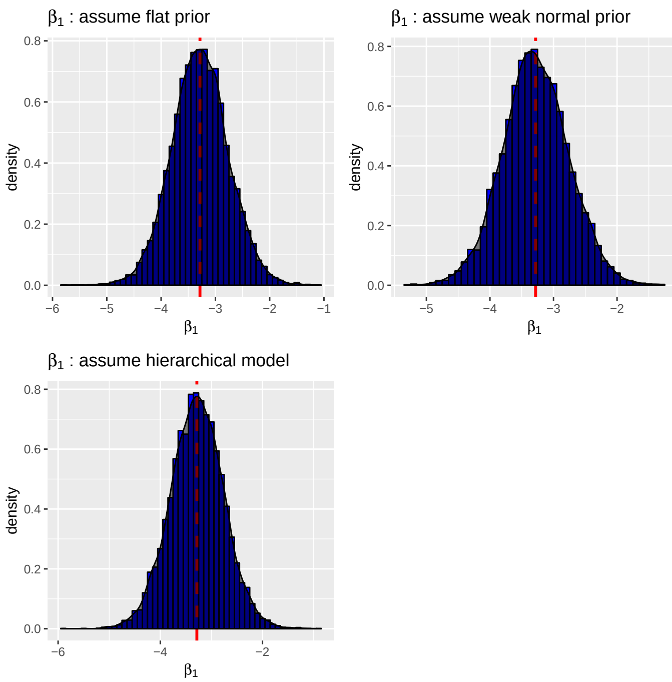
	Dbar	Dhat	pD	DIC
y	325.420	322.493	2.927	328.347
total	325.420	322.493	2.927	328.347

As we can see that all results are very close. Particularly from the DIC it suggests each of the above model is not necessarily superior than the other. So if we do want to pick one as our final model, we could use the most parsimonious one, and here definitely we do not need to bother with using hierarchical model.

Now for part (i):

Since the estimate of β_1 *is negative, we conclude that the number of people drown per year is declining. The following R code plot the histogram of* β_1 *(slope). The data is extracted from the coda generated in winBUGS.*

```r
#HW 5

#the csv file from the coda generated in winBUGS
data <- read.csv("C:\\akira\\data\\coda_flat_prior.csv", header = TRUE)
data1 <- read.csv("C:\\akira\\data\\coda_normal_prior.csv", header = TRUE)
data2 <- read.csv("C:\\akira\\data\\coda_hierarchical.csv", header = TRUE)

#head(data)

library(ggplot2)
library(gridExtra)

#histogram of the slope beta1

#flat prior
p_beta1 <- ggplot(data = data, aes(x = beta1, y = ..density..)) +
  geom_histogram(binwidth = 0.1, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(beta1)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black") +
  labs(title=expression(paste(beta[1],
    " : assume flat prior")), x = expression(beta[1]))

#normal weak prior
p1_beta1 <- ggplot(data = data1, aes(x = beta1, y = ..density..)) +
  geom_histogram(binwidth = 0.1, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(beta1)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black") +
  labs(title=expression(paste(beta[1],
  " : assume weak normal prior")), x = expression(beta[1]))

#hierarchical model
p2_beta1 <- ggplot(data = data2, aes(x = beta1, y = ..density..)) +
  geom_histogram(binwidth = 0.1, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(beta1)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black") +
  labs(title=expression(paste(beta[1],
    " : assume hierarchical model")), x = expression(beta[1]))

grid.arrange(p_beta1, p1_beta1, p2_beta1, nrow = 2)
```



The histogram further supported our conclusion, since not only the point estimate of β_1*, but also the whole distribution of* β_1 *lies on the negative values.*

For part (ii):

We can read from the point estimate above that, for all three different approaches, the predicted number of people drown in 2016 *is about* 156 *to* 157*.*

For part (iii):

The following R code plot the histogram of the posterior predictive distribution of y_p *for the year of* 2016*:*

```
#histogram of model predict for year 2016

#flat prior
p_yp <- ggplot(data = data, aes(x = y_p, y = ..density..)) +
  geom_histogram(binwidth = 10, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(y_p)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black")  +
  labs(title = expression(paste(y[p],
                  " 2016: flat prior")),
       x = expression(y[p]))

#normal weak prior

p1_yp <- ggplot(data = data1, aes(x = yp, y = ..density..)) +
```

```
  geom_histogram(binwidth = 10, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(yp)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black") +
  labs(title = expression(paste(y[p],
      " 2016: weak normal prior")),
       x = expression(y[p]))

#hierarchical model

p2_yp <- ggplot(data = data2, aes(x = yp, y = ..density..)) +
  geom_histogram(binwidth = 10, color = "black", fill = "blue") +
  geom_vline(aes(xintercept=mean(yp)),
             color="red", linetype="dashed", size=1)  +
  geom_density(alpha=.5, fill="black") +
  labs(title = expression(paste(y[p],
      " 2016: hierarchical model")),
       x = expression(y[p]))

grid.arrange(p_yp, p1_yp, p2_yp, nrow = 2)
```