Problem 1

(a) The code for Gibbs sampling algorithms is attached in the Appendix.

From the output below, we see that the effective sample sizes are all much larger than 1000.

```
> for(i in 1:m){cat(effectiveSize(THETA[,i]),"\n")}
4551.612
4968.519
5000
5465.115
4096.163
4625.084
5576.988
4776.939
> for(i in 1:3){cat(effectiveSize(SMT[,i]),"\n")}
4739.021
4177.568
3618.343
```

(b) We can approximate the posterior mean and 95% confidence interval of σ^2, μ, and τ^2 using the MCMC samples:

```
> apply(SMT,2,mean)
   sigma2        mu      tau2
1.010420  7.553082  5.659925
> apply(SMT,2,function(x) quantile(x,prob=c(0.025,0.975)))
          sigma2       mu       tau2
2.5%   0.8203852 5.992710   2.064836
97.5%  1.2387424 9.077109  14.550425
```

We then compare it to the 95% confidence intervals of prior distributions of σ^2, μ, and τ^2:

```
> #prior 95% CI
> sigma2.quantile.prior<-c(qigamma(0.025,nu0/2,nu0*s20/2),qigamma(0.975,
    nu0/2,nu0*s20/2))
> mu.quantile.prior<-qnorm(c(0.025,0.975),mu0,g20)
> tau2.quantile.prior<-c(qigamma(0.025,eta0/2,eta0*t20/2),qigamma(0.975,
    eta0/2,eta0*t20/2))
> sigma2.quantile.prior
[1]   4.066275 592.468353
> mu.quantile.prior
[1] -2.79982  16.79982
> tau2.quantile.prior
[1]   2.71085  394.97890
```

It is noticed that the posterior confidence interval in much narrower compared with the prior confidence interval. The 95% posterior confidence interval of σ^2 does not even overlap with the 95% prior confidence interval. The data provides us information to make more precise and accurate estimates of the parameters we are interested in.

(c) Figure 1 shows the approximate prior and posterior densities of R. The region closed to 1 has very high posterior density while the prior density of R is relatively flat. This suggest that the between-group variance is likely to be much larger than the within-group variance.

Notice that the both densities for $R > 0$ should be 0. The positive densities of R for $R > 0$ on the Figure 1 may be caused by the implementation of density() function.

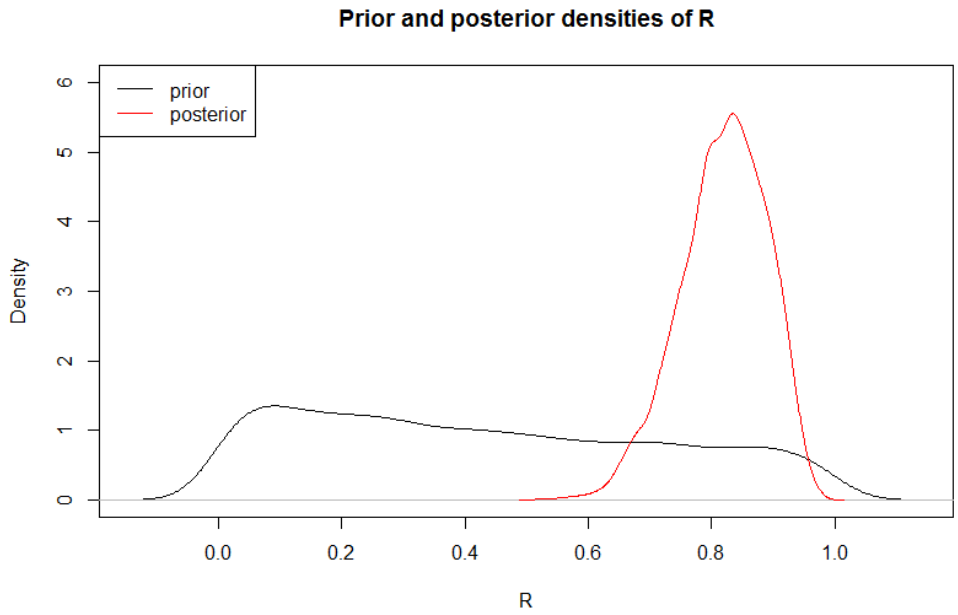

Figure 1: Prior and posterior density of R

(d) Using MCMC approximation, we see that the the posterior probability that θ_7 is smaller than θ_6 is 0.5818. The posterior probability that θ_7 is the smallest of all the θ's is 0.436.

```
> mean(THETA[,7]<THETA[,6])
[1] 0.5818
> #The posterior probability that theta7 is the smallest of all theta's
> mean(apply(THETA,1,min)==THETA[,7])
[1] 0.436
```

(e) Below is the plot showing the sample averages and posterior expectation of θ for each school. We can see that for all schools the sample averages are very closed to the posterior expectations of the mean parameters $\theta's$. One reason for this is that the sample sizes of the 8 groups do not differ much. Therefore, the shrinkage in this example is not that significant.

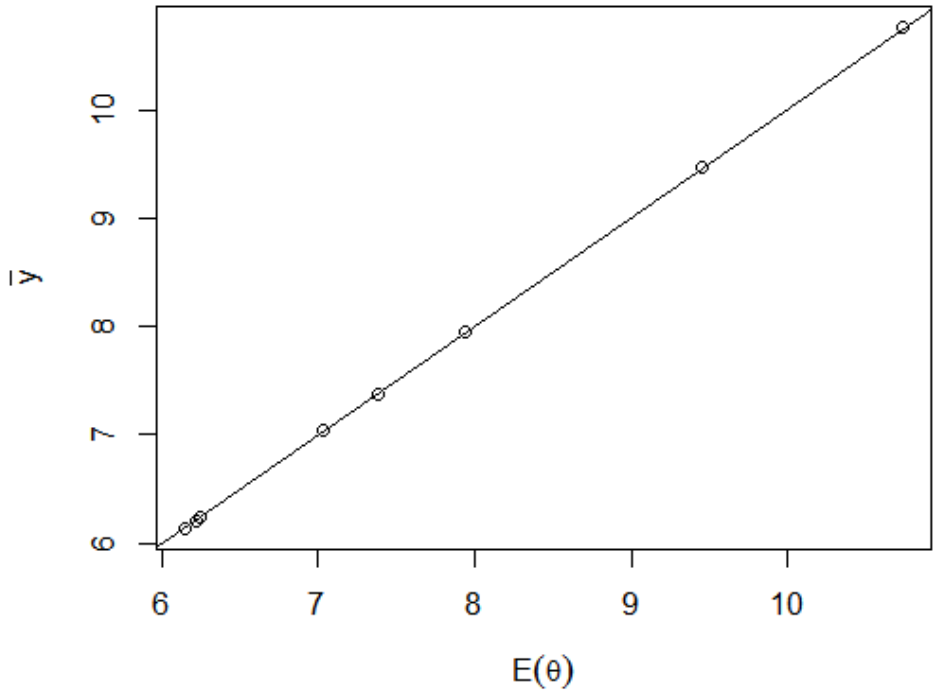

Figure 2: Sample averages $\bar{y}_1, ..., \bar{y}_8$ against the posterior expectations of $\theta_1, ..., \theta_8$

The output below shows that the sample mean of all observations is 7.6913, and the posterior mean of μ is 7.5531.

```
> sum(as.numeric(lapply(school,sum)))/sum(as.numeric(lapply(school,length))
    )
[1] 7.691278
> #posterior expectation of mu
> mean(SMT[,2])
[1] 7.553082
```

Problem 2

(a) As $\text{logit}[Pr(Y_i = 1 \mid \alpha, \beta, x_i)] = \alpha + \beta x_i$, we know that $Pr(Y_i = 1 \mid \alpha, \beta, x_i) = \text{logistic}(\alpha + \beta x_i)$. Hence we can see that:

$$Y_i \mid \alpha, \beta, x \sim \text{Bernoulli}(\text{logistic}(\alpha + \beta x_i))$$

Therefore,

$$\begin{aligned}\prod_{i=1}^{n} p(y_i \mid \alpha, \beta, x_i) &= \prod_{i=1}^{n} \left[Pr(Y_i = 1 \mid \alpha, \beta, x_i)^{y_i} \left(1 - Pr(Y_i = 1 \mid \alpha, \beta, x_i)\right)^{1-y_i} \right] \\ &= \prod_{i=1}^{n} \text{dbinom}\left(y_i, 1, \text{logistic}(\alpha + \beta)\right),\end{aligned}$$

where logistic is the inverse of the logit function.

(b) We let the regression coefficient to be i.i.d normal(0,100). The prior is convenient when implementing the algorithm and is used several times in Chapter 10 of the textbook. Here, α and β will be loosely centred at (0,0) and almost all possible values of $Pr(Y_i = 1 \mid \alpha, \beta, x_i)$ can be covered given the observed range of x.

(c) R-code to implement the Metropolis algorithm that approximates $p(\alpha, \beta \mid \mathbf{y}, \mathbf{x})$ is as follows:

```
n<-nrow(sparrow)
y<-sparrow[,1]
x<-cbind(rep(1,n),sparrow[,2])

#proposal distribution
beta<-c(logit(0.9)-(logit(0.9)-logit(0.1))*3,(logit(0.9)-logit(0.1))/5)
var.prop<-6*solve(t(x)%*%x)

#prior parameters
pmn.beta<-c(0,0)
psd.beta<-c(10,10)

S<-10000
acs<-0
BETA<-matrix(0,nrow = S,ncol = 2)
set.seed(1)

for(s in 1:S){
  beta.p<-t(rmvnorm(1,beta,var.prop))

  lhr<-sum(dbinom(y,1,logistic(x%*%beta.p),log = T))+
    sum(dnorm(beta.p,pmn.beta,psd.beta,log = T))-
    sum(dbinom(y,1,logistic(x%*%beta),log = T))-
    sum(dnorm(beta,pmn.beta,psd.beta,log = T))

  if(log(runif(1))<lhr){beta<-beta.p; acs<-acs+1}
  BETA[s,]<-beta
}
```

The acceptance rate is above 50% and the effective sample size for each regression coefficient is greater than 1000:

```
> effectiveSize(BETA)
    var1      var2
1090.755 1027.339
> acs/S
[1] 0.5105
```

(d) Figure 3 shows the prior and the posterior distribution of α, while Figure 4 is the plot of the prior and posterior distribution for β.

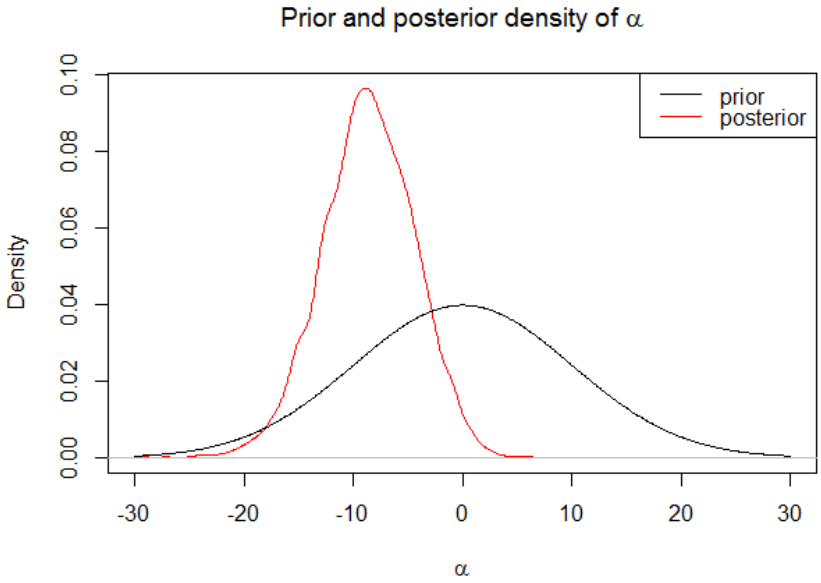

Figure 3: Prior and Posterior distribution of α

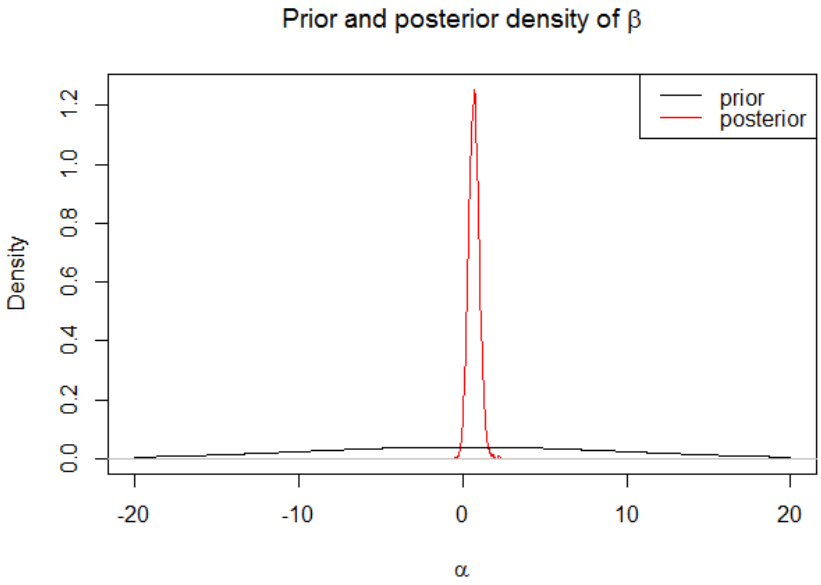

Figure 4: Prior and Posterior distribution of β

The posterior variances of both α and β are smaller than the prior variance. It is also noticed that the posterior variance of β is significantly smaller than that of α.

(e) For each x, we can compute the corresponding values of f for each (α, β) in MCMC draws, and then use the posterior quantile to approximate the band. Figure 5 is a plot of the confidence band:

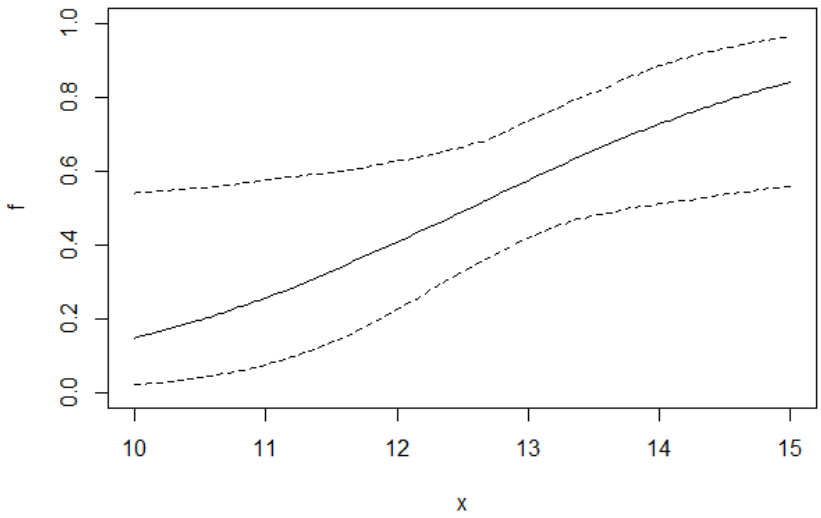

Figure 5: Confidence band of f

Problem 3

(a) We will let time and pH be the predictors of the linear regression. The quadratic term and interaction term are found to be not significant and thus is not included in the model. The ordinary least squares estimate of the coefficients are as follows:

```
> x<-cbind(rep(1,length(height)),time,ph)
> solve(t(x)%*%x)%*%t(x)%*%height #ols
             [,1]
      7.2086931
time  3.9910000
ph    0.5777521
```

The estimate of the intercept is 7.2087. It does not have any direct interpretation as zero is outside the observed range of pH values.

The estimated coefficient of time is 3.991. It suggests that the average height of tomato plants is about 3.991 units higher in the second measurement than in the first measurement if pH remains constant.

The estimated coefficient of pH is 0.5777, which means that the predicted height increases by 0.5777 unit when pH increases by 1 in the same measurement.

(b) Figure 6 is a plot of residuals versus index. We observe that the residuals are very likely to be correlated, as the residuals for the same plant in two measurements are relatively close. The assumptions of independence seems to be violated in the fitted model.

The coefficient of time is believed to be sensitive to this violation of assumptions.

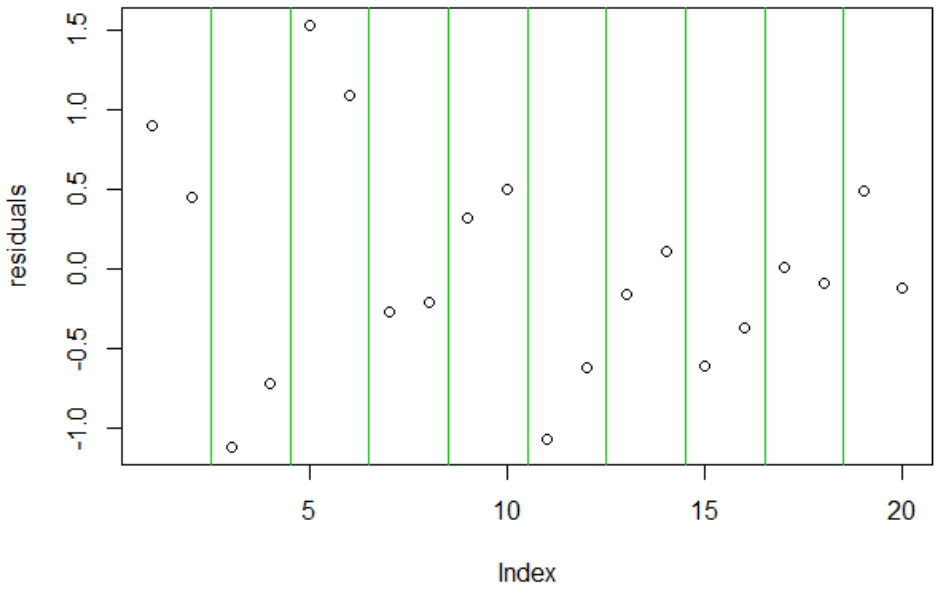

Figure 6: Residuals versus index plot

(c) Let Y_j denote the height of j^{th} observation. For i = 1,...,10, Y_{2i-1} and Y_{2i} are the heights of two measurements from the same tomato plant. We assume the observations within a plant are correlated and each pair of observations from different tomato plants are independent of each other. Hence we can use the following covariance matrix:

$$\Sigma = \sigma^2 \mathbf{C}_\rho = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \rho & \cdots & 0 \\ 0 & 0 & \rho & 1 & & \\ \vdots & \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & & 1 \end{pmatrix}$$

After running the MCMC algorithm, we apply thinning to the output and get 1000 posterior samples with low autocorrelation. The effective sample size is 875 for ρ, and 1000 for all other parameters. From the effective sample size and the ACF plots for parameters in Figure 7, we see that there are no significant problems.

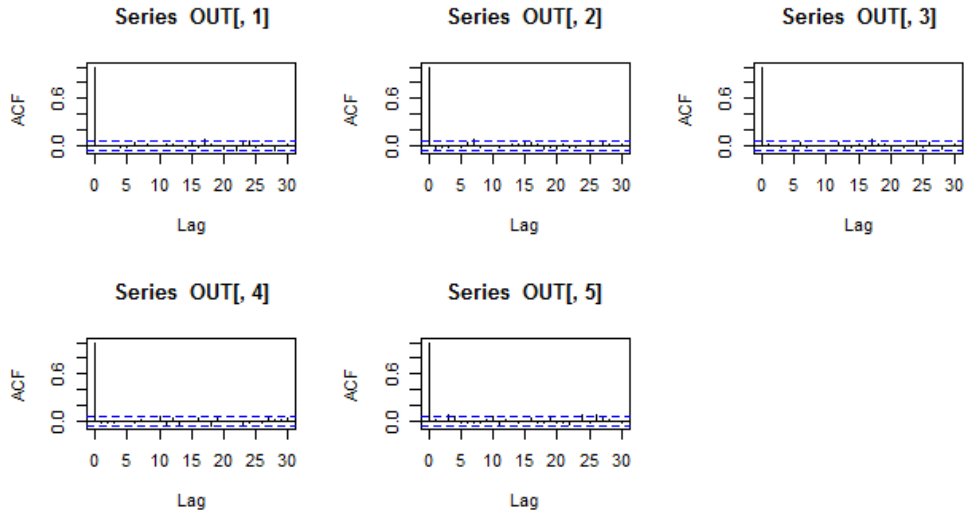

Figure 7: ACF plot for $\boldsymbol{\beta}$, σ^2, and ρ

(d) We can obtain the posterior expectation of the parameters by calculating the mean of MCMC posterior samples. In the output below, we see that the posterior mean of regression coefficients are similar to the OLS estimates. However, it is noticed that the posterior mean of residual variance is larger than the estimated residual variance of the linear regression model:

```
> apply(OUT,2,mean)
[1] 7.1869418 3.9941098 0.5784678 0.6442605 0.7967848
> coef(lmfit)
         x      xtime        xph
7.2086931 3.9910000 0.5777521
> summary(lmfit)$sigma^2
[1] 0.5379304
```

We also notice that the standard deviation of the coefficient for time is smaller(0.16) when we allow the observations within a plant to be correlated, compared to that in the OLS(0.33). The standard deviations for the intercept and the coefficient of pH are larger in the result of part(c) than in the linear regression model:

```
> apply(OUT,2,sd)[1:3]
[1] 0.8438835 0.1584024 0.1786438
> summary(lmfit)$coefficient[,"Std. Error"]
         x      xtime        xph
0.5890540 0.3280032 0.1203538
```

We would prefer the model used in part(c) as the posterior mean of ρ is 0.8, and 0 is not in the 95% confidence interval of ρ:

```
> mean(OUT[,5])
[1] 0.7967848
> quantile(OUT[,5],c(0.25,0.95))
      25%       95%
0.7406251 0.9388984
```