Exercise 3.1

In this section, $Z = X\beta + \epsilon$, $\epsilon \sim N(0, I)$. Using standard linear model results, if a priori the distribution of β is diffuse, then
$\beta|y, Z$ is distributed $N_k(\beta_z, (X'X)^{-1})$, where $\beta_z = (X'X)^{-1}X'Z$.
$Z_i|y, \beta \sim N(x_i^T\beta, 1)$ truncated at the left by 0 if $y_i = 1$.
$Z_i|y, \beta \sim N(x_i^T\beta, 1)$ truncated at the right by 0 if $y_i = 0$.

In practice it is customary to assign a flat noninformative prior to β.
Response: class variable (1/0).
Predictors: blood pressure, bmi and age.
The length of the MCMC chain was 5000.

Predictors	Estimates from Bayesian analysis	Estimates from GLM method
Blood pressure	-0.11	-0.17
BMI	0.43	0.71
Age	0.32	0.53

Traceplots show good mixing. All the plots are attached in back.

Exercise 3.2

y_i: Yes/No.
x_i: Age
$\text{logit}(p_i) = \beta x_i$
Let $\beta \sim N(0, \sigma^2)$
Let $\text{P}(y_i = 1|x_i, \beta) = h_\beta(x_i)$
$\text{P}(y_i = 0|x_i, \beta) = 1 - h_\beta(x_i)$
$\text{P}(y|\beta, x_i) = \prod(h_\beta(x_i))^{y_i}(1 - h_\beta(x_i))^{1-y_i}$
$\text{P}(\beta|y, x) \propto \text{P}(y|\beta, x)P(\beta)$
$\log \text{P}(\beta|y, x) = \sum y_i \log h_\beta(x_i) + \sum(1 - y_i)\log(1 - h_\beta(x_i)) - \frac{1}{2\sigma^2}\beta^2 + K$
Using the lemma $\frac{\partial}{\partial\beta}h_\beta(x_i) = -h_\beta(x_i)(1 - h_\beta(x_i))x_i$ we get:

$\frac{\partial}{\partial\beta}l(\beta) = \sum(y_i - h_\beta(x_i))x_i - \frac{\beta}{\sigma^2}$

Exercise 3.3

By GLM method I got $\hat{\beta} = -0.1211981 \approx -0.121$.
By Gradient descent method for 10,000 iterations, I got $\hat{\beta} = -0.128117 \approx -0.128$.

Both methods give almost same estimate. Figure 3 shows that the estimate becomes close to GLM estimate as the number of iterations increases.

Exercise 3.4

The mean of the Gaussian is the MAP estimate of the log posterior and the precision is just the negative of the Hessian matrix.
lemma: $\frac{\partial}{\partial \beta} h_\beta(x_i) = -h_\beta(x_i)(1-h_\beta(x_i))x_i$
$\log p(\beta|y,x,\lambda) = \sum y_i \log h_\beta(x_i) + (1-y_i)\log(1-h_\beta(x_i)) - \frac{\lambda}{2}\beta^2$
$\frac{\partial}{\partial \beta} \log p(\beta|y,x,\lambda) = \sum (y_i - h_\beta(x_i))x_i - \lambda\beta$
$\frac{\partial^2}{\partial \beta^2} \log p(\beta|y,x,\lambda) = -\sum (y_i - h_\beta(x_i))x_i^2 - \lambda.$

Exercise 3.5

$\log p(\beta|y,x,\lambda) = y^T \log h(X\beta) + (1-y)^T \log(1-h(X\beta)) - \frac{\lambda}{2}\beta^T\beta$
$\frac{\partial}{\partial \beta} \log p(\beta|y,x,\lambda) = X^T(y-h(X\beta)) - \lambda\beta$
$\frac{\partial^2}{\partial \beta^2} \log p(\beta|y,x,\lambda) = -X^T\text{diag}(h(X\beta)(1-h(X\beta)))X - \lambda = H.$
$\beta|y,x,\lambda \sim$ multivariate gaussian with mean β_{MAP} and covariance matrix $-H$ (negative of hessian matrix).
So maximum a posteriori estimate and 95% marginal credible intervals are:

parameter	Estimate	Credible intervals
Intercept	-0.35	(-0.49, -0.20)
Age	-0.12	(-0.27, 0.02)

Exercise 3.6

Counts may exhibit larger variance than expected under the Poisson model. This is referred to as over-dispersion. This over-dispersion may be due to random behavior of the mean or parameter, i.e. when the parameter is a random variable. However if the mean follows a gamma distribution, then the counts follow a negative binomial distribution. This allows for the variance to be proportional to the mean.
Having said that, a normal distribution is often a rather good approximation to a Poisson one for data with a mean above 30 or so. And in a regression framework, where we have predictors influencing the count, an OLS with its normal distribution may be easier to fit and would actually be more general, since the Poisson distribution and regression assume that the mean and the variance are equal, while OLS can deal with unequal means and variances -

for a count data model with different means and variances, one could use a negative binomial distribution, for instance.

Exercise 3.7

Using Grade^2 as the only covariate, we get the following results:

parameter	Estimate	Credible intervals
Intercept	2.704	(2.699, 2.709)
Grade^2	0.073	(0.068, 0.079)

Exercise 3.9

Traceplots are given in "Plots" section.

parameter	Estimate	Credible intervals
Intercept	2.38	(2.36, 2.39)
Grade	0.04	(0.04, 0.05)

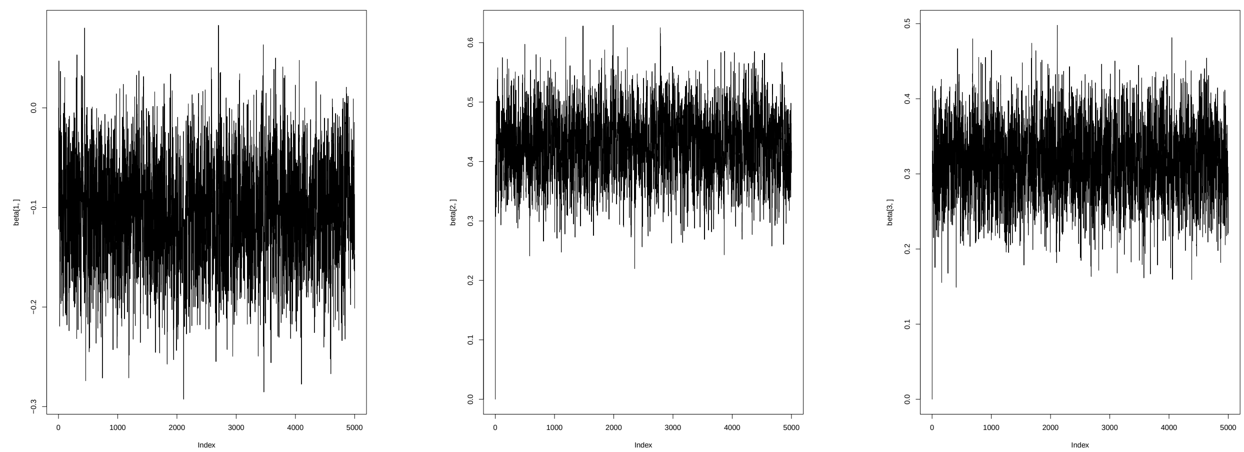
Exercise 3.10

More covariates have been added to the model. Traceplots are given in "Plots" section.

parameter	Estimate	Credible intervals
Intercept	-0.53	(-1.77, 0.68)
Grade	0.11	(0.03, 0.19)
Male	0.36	(0.11, 0.63)
Black	2.11	(1.47, 2.76)
Hispanic	2.13	(1.50, 2.78)
White	1.52	(1.24, 1.82)
Attendance	-0.62	(-1.09, -0.16)
Race	1.43	(0.29, 2.53)

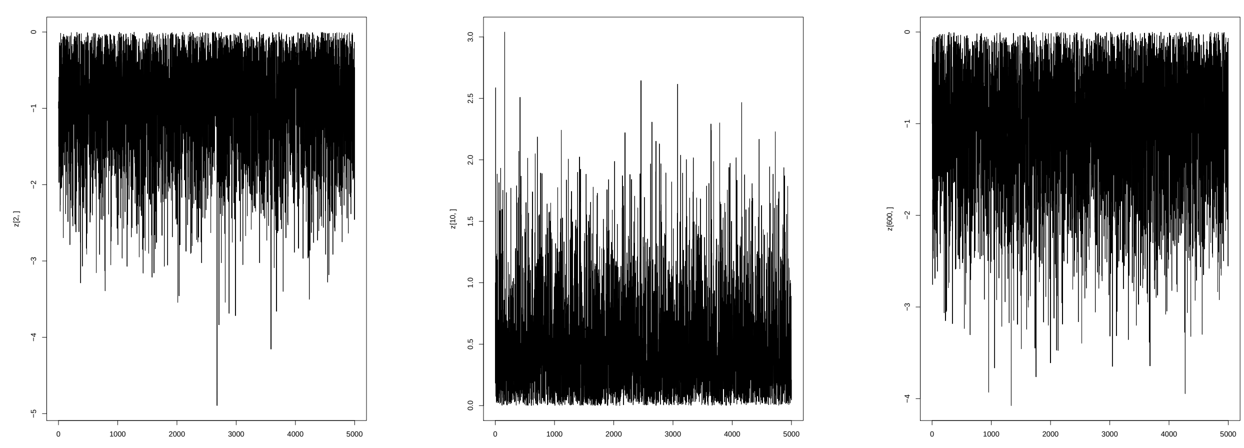
Exercise 3.11

Here we can either incorporate a binary latent variable or impute the data.
Let $z_i = 1$ if $y_i \geq 4$ or 0 otherwise.



(a) Trace plot for β_1 (b) Trace plot for β_2 (c) Trace plot for β_3

Figure 1: 3.1 graphs



(a) Trace plot for Z_2 (b) Trace plot for Z_{10} (c) Trace plot for Z_{600}

Figure 2: 3.1 graphs

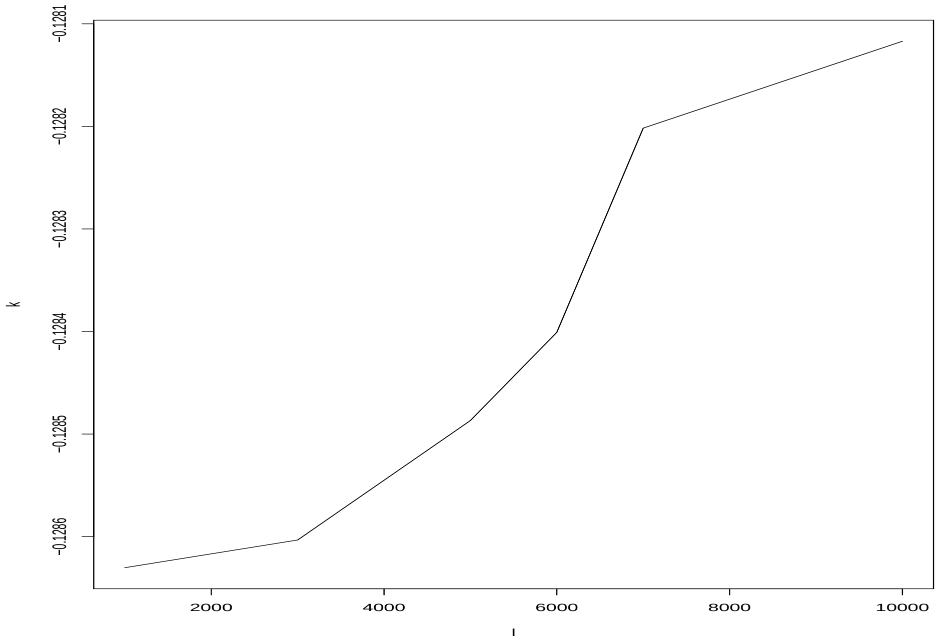


Figure 3: 3.3 β estimate with number of iterations

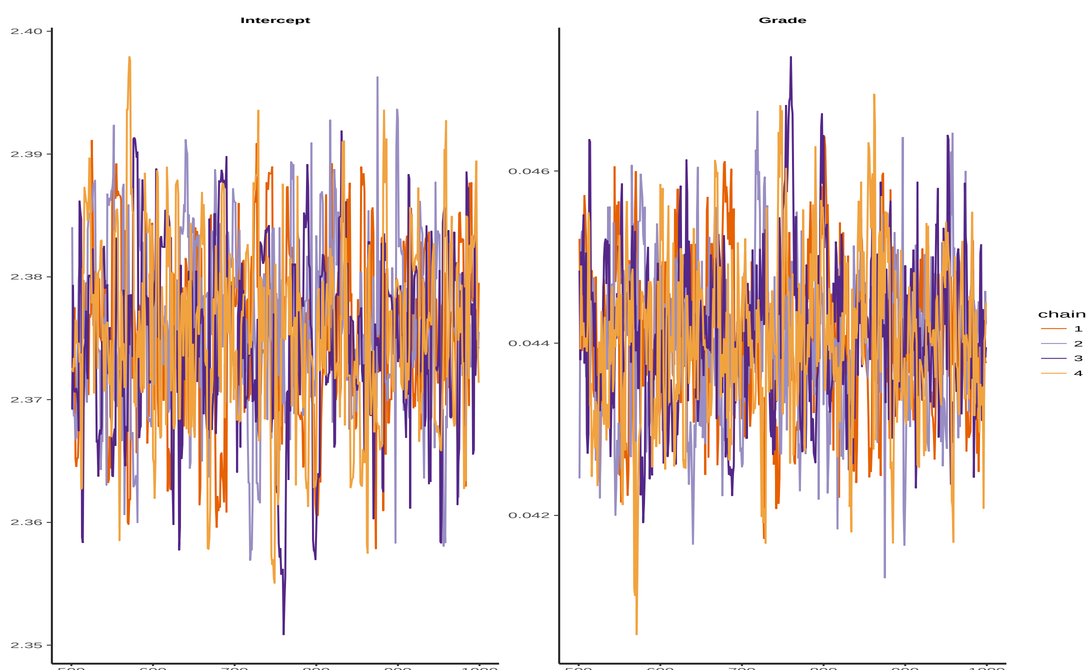


Figure 4: 3.9 traceplots of Intercept and Grade

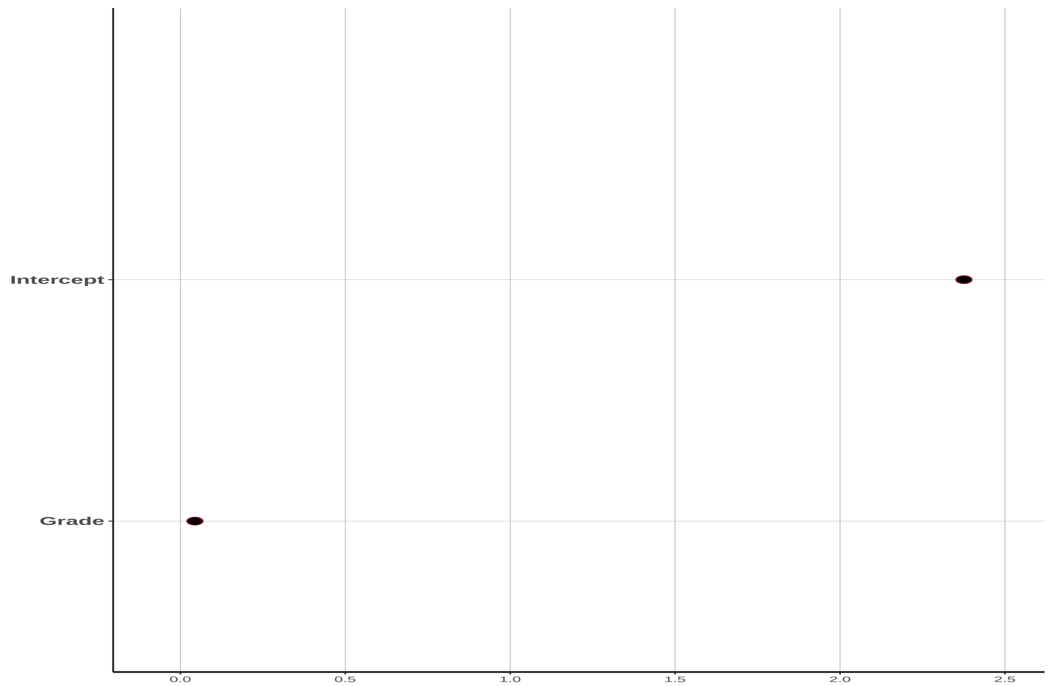


Figure 5: 3.9 Credible intervals

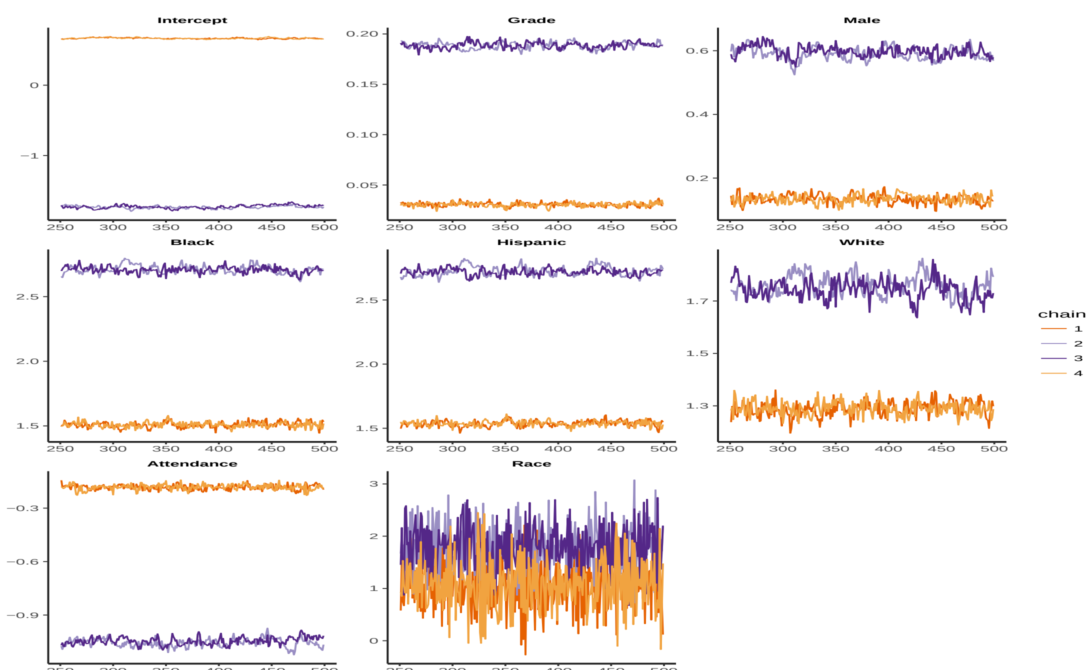


Figure 6: 3.10 traceplots

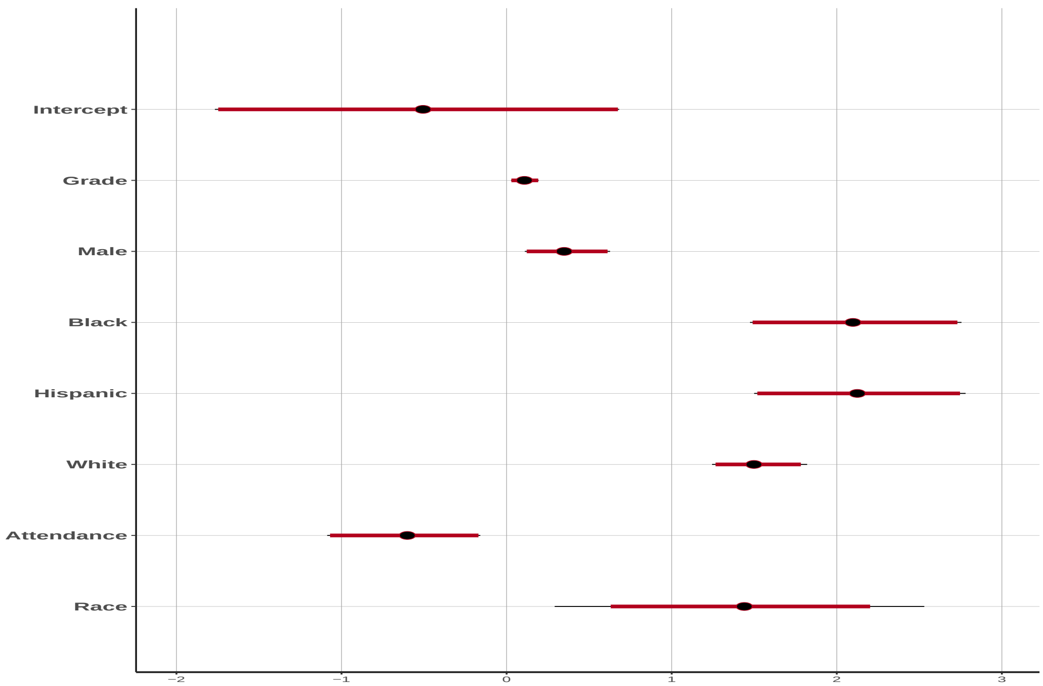


Figure 7: 3.10 Intervals