Assignment 7

AUTHOR
Joaquin Peraza

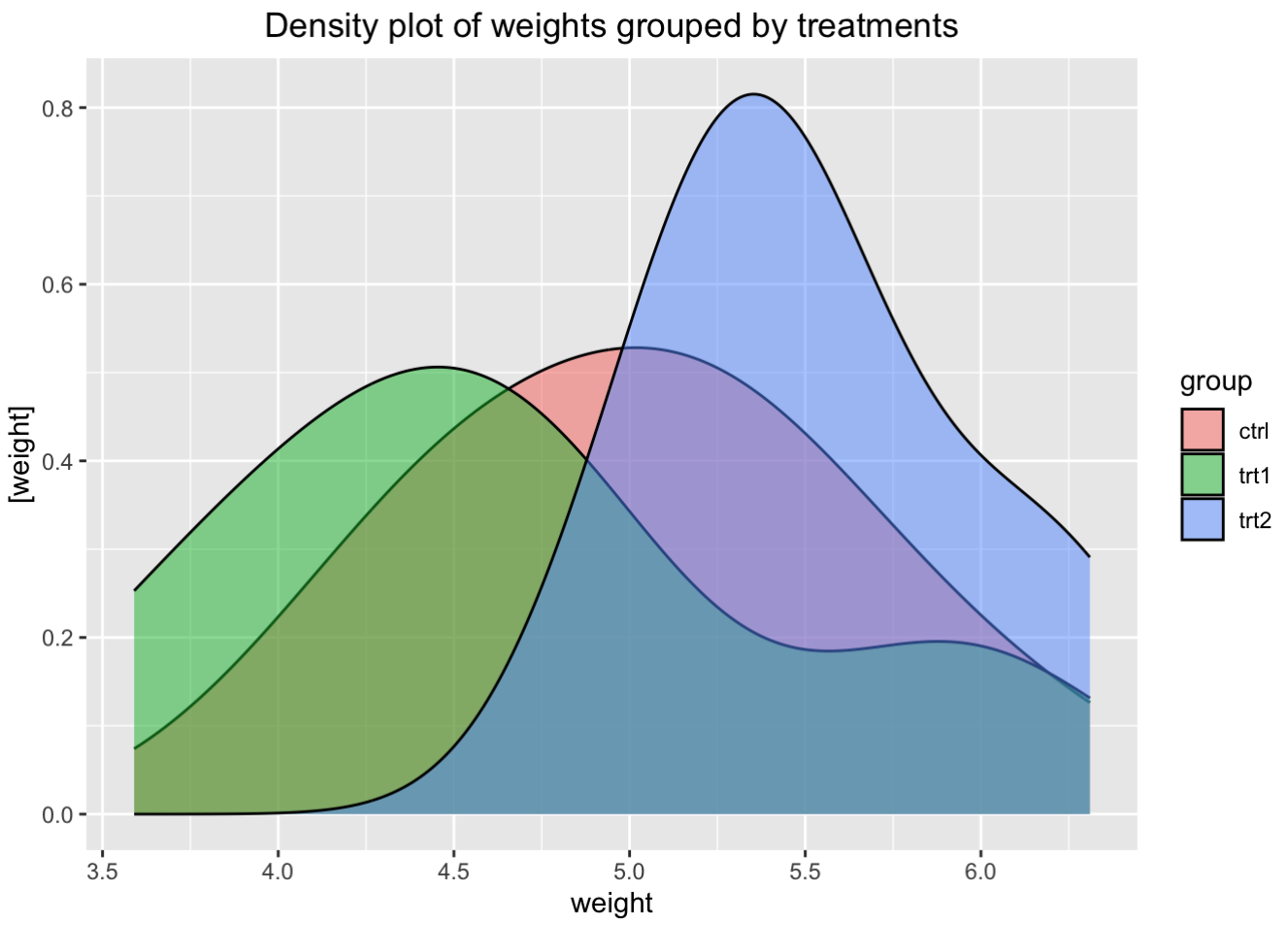
1.

```
set.seed(0)

library(ggplot2)

library(faraway)
knitr::kable(list(PlantGrowth[1:10,], PlantGrowth[10:19,],
                  PlantGrowth[20:29,]), row.names = TRUE)
```

	weight	group		weight	group		weight	group
1	4.17	ctrl	10	5.14	ctrl	20	4.69	trt1
2	5.58	ctrl	11	4.81	trt1	21	6.31	trt2
3	5.18	ctrl	12	4.17	trt1	22	5.12	trt2
4	6.11	ctrl	13	4.41	trt1	23	5.54	trt2
5	4.50	ctrl	14	3.59	trt1	24	5.50	trt2
6	4.61	ctrl	15	5.87	trt1	25	5.37	trt2
7	5.17	ctrl	16	3.83	trt1	26	5.29	trt2
8	4.53	ctrl	17	6.03	trt1	27	4.92	trt2
9	5.33	ctrl	18	4.89	trt1	28	6.15	trt2
10	5.14	ctrl	19	4.32	trt1	29	5.80	trt2

The dataset PlantGrowth contains observations of the growth of plants under three different treatments. The treatments are indicated by the variables “ctrl”, “trt1”, and “trt2” in the column “group”. There are 10 observations for each treatment, resulting in a total of 30 observations. The response variable is the dry weight of the plants, measured in grams. The “ctrl” treatment represents a control group where the plants were grown under normal conditions, while the “trt1” and “trt2” treatments represent groups where the plants were subjected to different experimental treatments. The specific details of the treatments are not provided in the dataset.

```
ggplot(data = PlantGrowth, aes(x = weight, fill=group,group=group)) +
  geom_density(adjust=1.5, alpha=0.5) + xlab("weight") + ylab("[weight]") +
  ggtitle("Density plot of weights grouped by treatments") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
knitr::kable(aggregate(by=list(PlantGrowth$group),
            list(Weight=PlantGrowth$weight), FUN=mean),
             caption = "Mean weight by treatment")
```

Mean weight by treatment

Group.1	Weight
ctrl	5.032
trt1	4.661
trt2	5.526

2. Design matrix

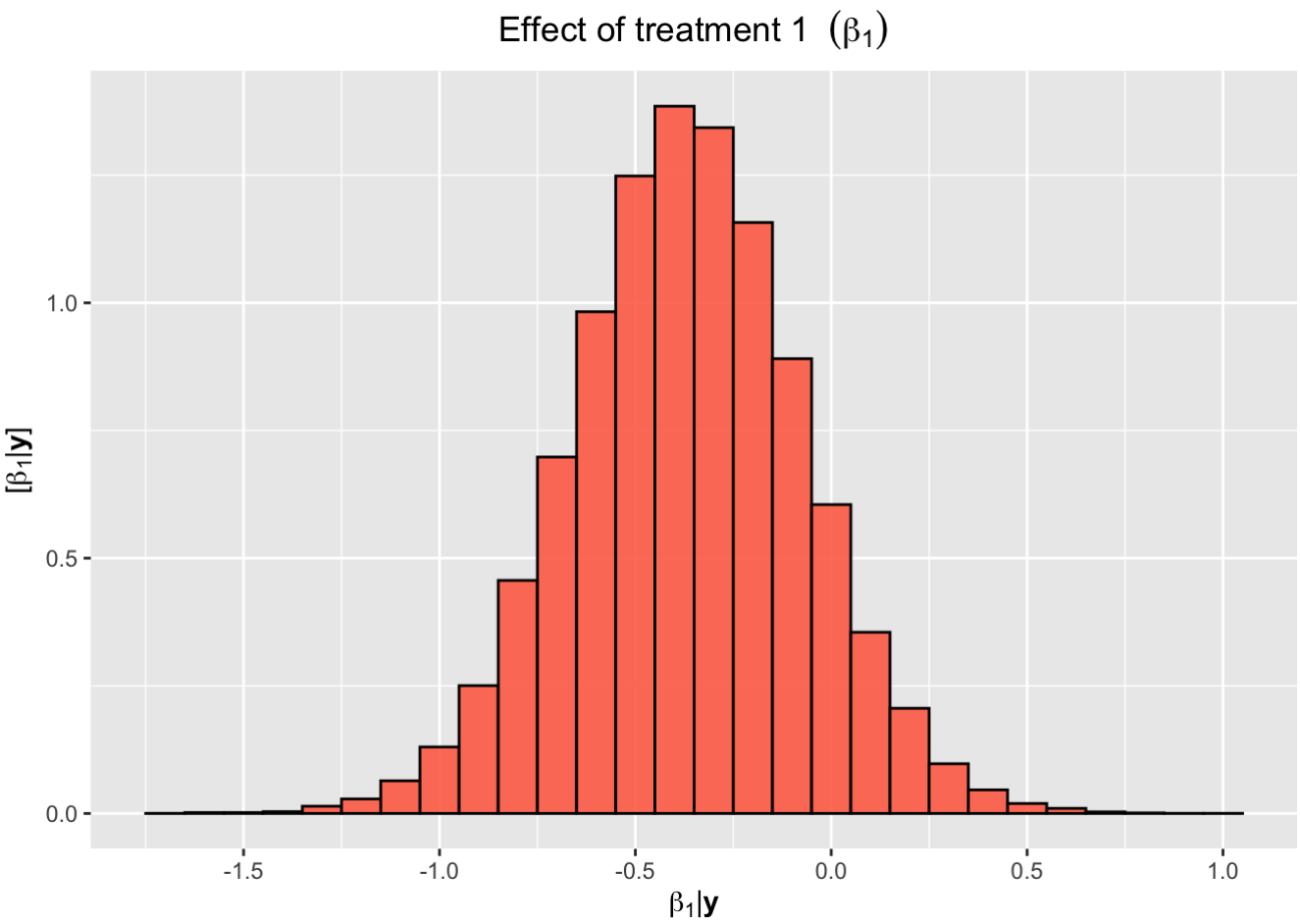
```
#Design matrix
X <- model.matrix(~ group, data = PlantGrowth)
```

3-4. Gibbs sampler for the linear model

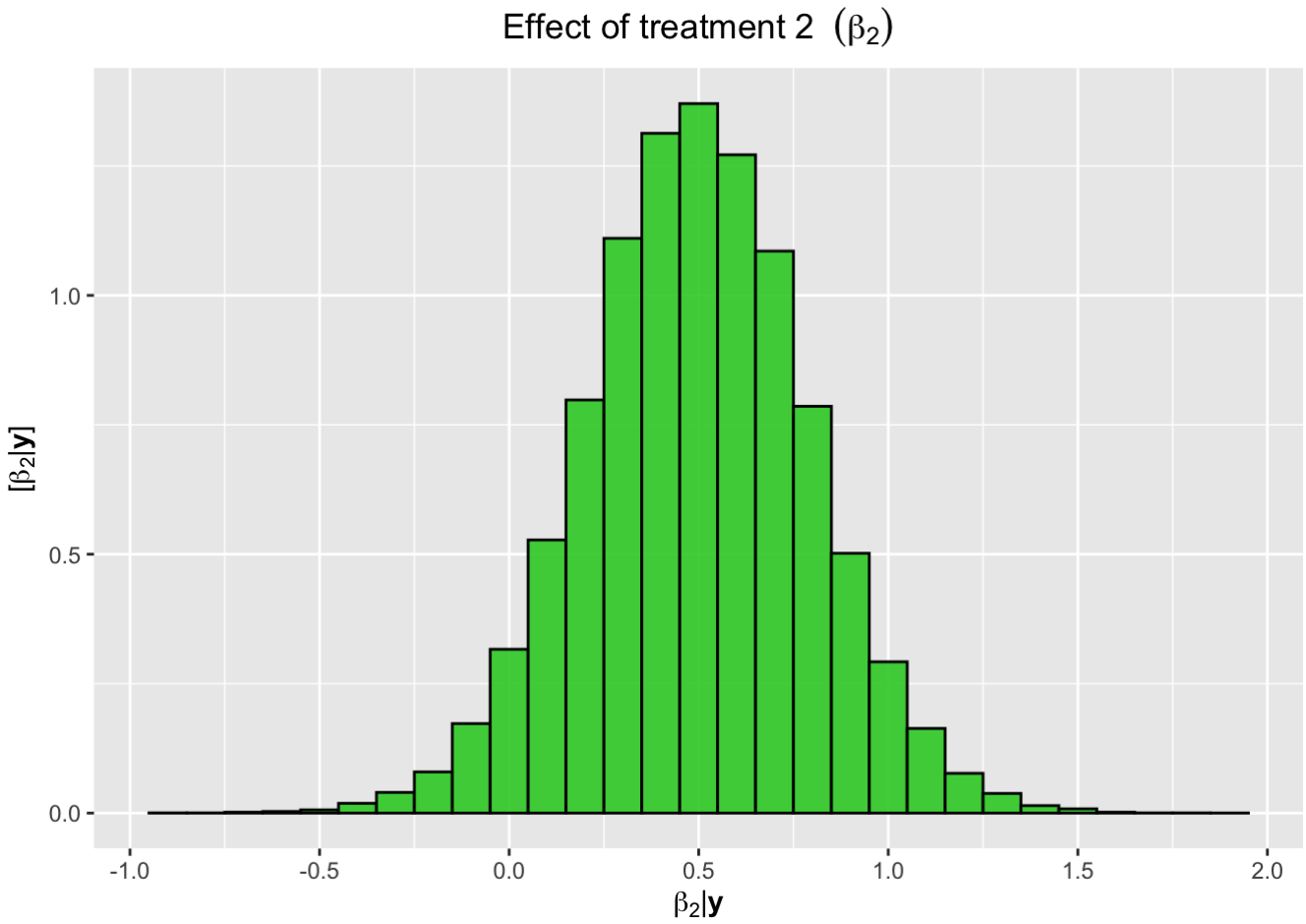
```r
# Preliminary steps
# Required package to sample from multivariate normal distribution
library(mvnfast)
# Response variable
y <- PlantGrowth$weight

n <- length(y) # Number of observations
p <- 3 # Dimensions of beta

K <- 50000  #Number of MCMC samples to draw

samples <- as.data.frame(matrix(,K,4)) # Samples of the parameters we will save
colnames(samples) <- c(colnames(X),"sigma2")

sigma2.e <- 1  # Starting values for sigma2

sigma2.beta <- 10^3 #Prior variance for beta
q <- 2 #Inverse gamma prior with E() = 1/(r*(q-1)) and
r <- 1 #Var()= 1/(r^2*(q-1)^2*(q-2))



# MCMC algorithm
for(k in 1:K){
  # Sample beta from full-conditional distribution
  H <- solve(t(X)%*%X + diag(1/sigma2.beta,p))
  h <- t(X)%*%y
  beta <-  t(rmvn(1,H%*%h,sigma2.e*H))

  # Sample sigma from full-conditional distribution
  q.temp <- q+n/2
  r.temp <- (1/r+t(y-X%*%beta)%*%(y-X%*%beta)/2)^-1
  sigma2.e <- 1/rgamma(1,q.temp,,r.temp)

  # Save samples of beta and sigma
  samples[k,] <- c(beta,sigma2.e)
}


burn.in <- 5000
```

Posterior distributions for treatment 1 and 2 effects.

```
par(mfrow=c(2,1),mar=c(5,6,1,1))


ggplot() + aes(samples[-c(1:burn.in),2]) +
geom_histogram( binwidth=0.1, fill="coral1", color="black", alpha=0.9,
aes(y=after_stat(density))) + ggtitle(expression("Effect of treatment 1 "~(beta[1]))) +
xlab(expression(beta[1]*"|"*bold(y))) + ylab(expression("["*beta[1]*"|"*bold(y)*"]")) +
scale_x_continuous(breaks = seq(-2, 2, by = 0.5)) + scale_y_continuous(n.breaks = 5) +
theme(plot.title = element_text(hjust = 0.5))
```

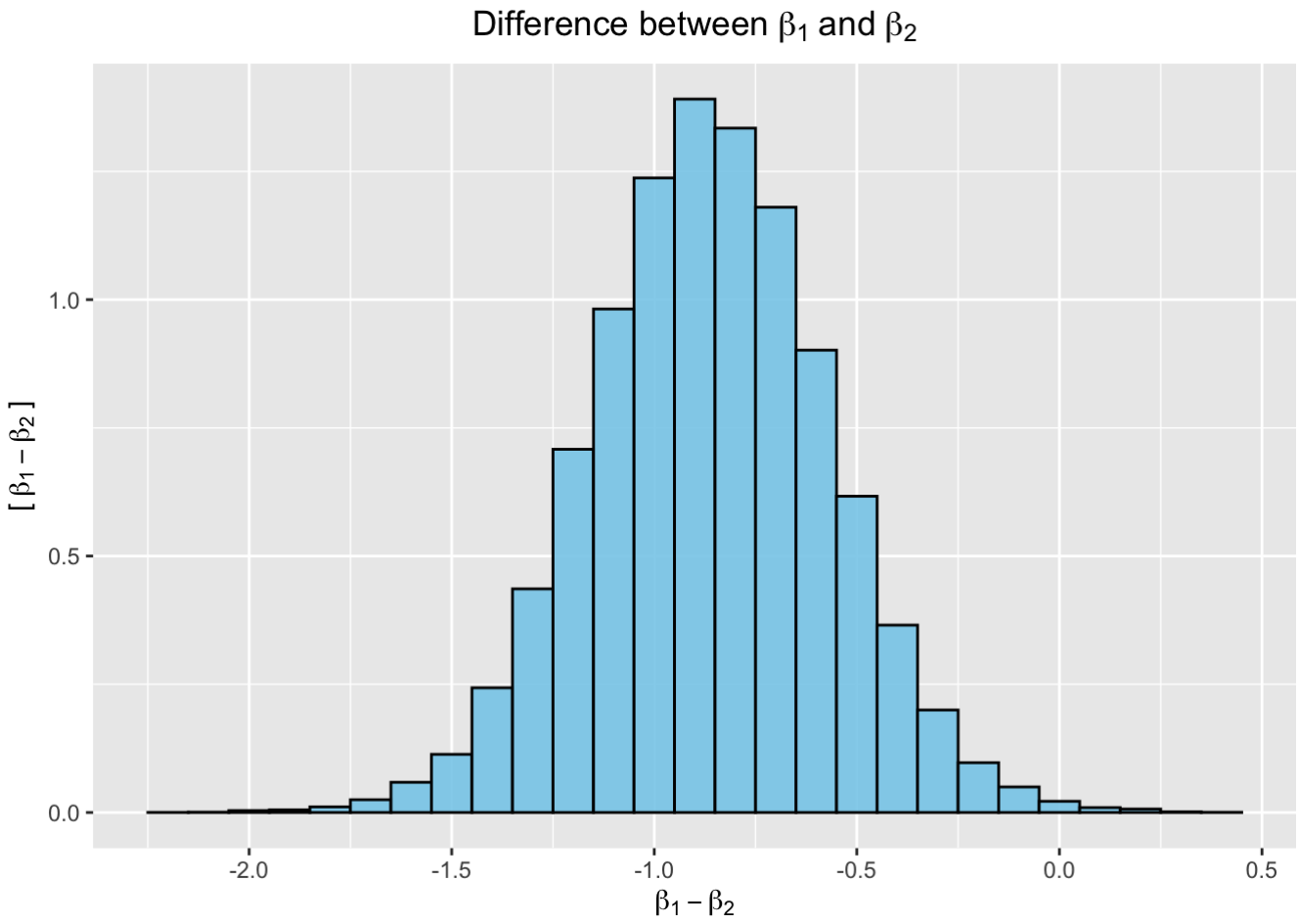



```
ggplot() + aes(samples[-c(1:burn.in),3]) +
geom_histogram( binwidth=0.1, fill="limegreen", color="black", alpha=0.9,
aes(y=after_stat(density))) + ggtitle(expression("Effect of treatment 2 "~(beta[2]))) +
xlab(expression(beta[2]*"|"*bold(y))) + ylab(expression("["*beta[2]*"|"*bold(y)*"]")) +
scale_x_continuous(breaks = seq(-2, 2, by = 0.5)) + scale_y_continuous(n.breaks = 5) +
theme(plot.title = element_text(hjust = 0.5))
```


Effect of treatment 2 (β_2)
1.0
0.5
0.0
$[\beta_2|\mathbf{y}]$
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
$\beta_2|\mathbf{y}$

5. Posterior distribution for the different between treatment 1 and 2.

```
# Extract beta_1 and beta_2 samples
beta_samples <- samples[-c(1:burn.in), 2:3]

# Calculate the difference between samples
beta_diff <- beta_samples[,1] - beta_samples[,2]

ggplot() + aes(beta_diff) +
geom_histogram( binwidth=0.1, fill="skyblue", color="black", alpha=0.9,
aes(y=after_stat(density))) +
ggtitle(expression("Difference between"~beta[1]~"and"~beta[2])) +
xlab(expression(beta[1]-beta[2])) + ylab(expression("["~beta[1]-beta[2]~"]")) +
scale_x_continuous(breaks = seq(-2, 2, by = 0.5)) + scale_y_continuous(n.breaks = 5) +
theme(plot.title = element_text(hjust = 0.5))
```




```
# Calcualte the Highest Density Interval
HDInterval::hdi(beta_diff, credMass = 0.95)
```

```
     lower      upper
-1.4271067 -0.2858243
```

```
attr(,"credMass")
[1] 0.95
```

```r
# Calculate the probability of beta_diff being positive
sum(beta_diff > 0) / length(beta_diff)
```

```
[1] 0.002933333
```

6.

The posterior distribution of the difference between treatment 1 and 2 appears to be skewed to the left, as shown in the histogram plot of `beta_diff`, which suggests that treatment 1 has a lower effect than treatment 2 on the plant growth. The x-axis represents the difference between the beta coefficients (treatment effects), and the y-axis represents the density of the posterior distribution. The 95% Highest Density Interval (HDI) for the posterior distribution of the differences between treatment 1 and 2 is [-1.4271, -0.2858], which means that there is a 95% probability that the true difference between the beta coefficients of treatment 1 and 2 falls within this interval. This interval confirms the previous conclusion that treatment 2 is likely to grow more than treatment 1. The proportion of `beta_diff` values that are greater than 0 is estimated to be 0.00293, which indicates that there is a very low probability of the difference between treatment 1 and 2 being greater than 0. This further supports the conclusion that treatment 2 effects is bigger than treatment 1.