Survmeth 895 Homework 4

David (Daiwei) Zhang

February 12, 2017

Problem 1(a)

Here the model is $Y_i \sim N(\mu, \sigma^2)$. We use the fact that $\frac{(n-1)s^2}{\sigma^2}|y \sim \chi^2_{n-1}$ and $\mu|\sigma^2, y \sim N(\bar{y}, \sigma^2/n)$ to simulate the posterior and predict the unobserved variables.

```
dat <- read.csv("survmeth895_hw4_dataset.csv", header = TRUE)
dat.inc <- na.omit(dat)
n <- 10
m <- 24
x <- dat.inc$Total.Cultivated.Area
y <- dat.inc$Area.Under.Wheat
s <- sd(y)

n.sim <- 1e3
sigma.sq <- (n-1) * s^2 / rchisq(n.sim, n-1)
mu <- rnorm(n.sim, mean(y), sqrt(sigma.sq / n))
y.pred <- rnorm(n.sim * m, mu, sqrt(sigma.sq))
y.pred <- matrix(y.pred, n.sim, m)
total.wheat <- rowSums(y.pred) + sum(y)

mean(total.wheat)
```

```
## [1] 4931.85
```

```
quantile(total.wheat, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 2572.159 7194.269
```

Problem 1(b)

Here the model is $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$. We use the fact that $\frac{(n-k)s^2}{\sigma^2}|y \sim \chi^2_{n-k}$ and $\beta|\sigma^2, y \sim N(\hat{\beta}, V_\beta \sigma^2)$, where $\hat{\beta} = (X^T X)^{-1}(X^T Y)$ and $V_\beta = (X^T X)^{-1}$ to simulate the posterior and predict the unobserved variables.

```
dat.inc <- na.omit(dat)
dat.exc <- dat[rowSums(is.na(dat)) > 0,]
n <- 10
x <- dat.inc$Total.Cultivated.Area
x <- cbind(rep(1, n), x)
x.exc <- dat.exc$Total.Cultivated.Area
x.exc.n <- length(x.exc)
y <- dat.inc$Area.Under.Wheat
b.hat <- solve(t(x) %*% x) %*% (t(x) %*% y)
b.hat <- c(b.hat)
b.var <- solve(t(x) %*% x)
```

```
s <- sd(y)

sigma.sq <- (n-2) * s^2 / rchisq(n.sim * x.exc.n, n - 2)
sigma.sq <- matrix(sigma.sq, n.sim, x.exc.n)
total.wheat <- rep(sum(y), n.sim)
for (i in 1 : n.sim){
    for (j in 1:x.exc.n){
        b <- mvrnorm(1, b.hat, b.var * sigma.sq[i, j])
        mu <- b[1] + b[2] * x.exc[j]
        y.pred <- rnorm(1, mu, sqrt(sigma.sq[i,j]))
        total.wheat[i] = total.wheat[i] + y.pred
    }
}

print(mean(total.wheat))
```

```
## [1] 5002.174
```

```
print(quantile(total.wheat, c(0.025, 0.975)))
```

```
##    2.5%   97.5%
## 3697.11 6219.68
```

Problem 1(c)

By comparing the results in 1(b) and 1(c), we see that the information on the total cultivated area is crucial in predicting to total area under wheat, as seen in the difference in the predicted mean and the width of the credibility intervals.