Homework 7, STA 360

Austin Kao

Hoff 7.3

```r
### Call necessary R libraries
library(mvtnorm)
### Load in the data
data_bluecrab <- read.table("data/bluecrab.dat", header = FALSE)
data_orangecrab <- read.table("data/orangecrab.dat", header = FALSE)
```

    a.

```r
### Generate posterior samples for blue crabs
### Code is based off Olivier's provided lab code
Y = data_bluecrab
### Set a seed for reproducibility
set.seed(12345)
### Set initial values and define certain quantities
mu0 = colMeans(Y)
S0 = cov(Y)
nu0 = 4
Sigma0 = 100*diag(c(1,1))
Y.sum = colSums(Y)
n = nrow(Y)
nu.n = n + nu0
### Set number of samples to generate
n.iter = 10000
### Initialize empty arrays for theta and Sigma
theta = array(dim = c(n.iter, 2))
Sigma = array(dim = c(2, 2, n.iter))
### Add initial values to arrays
theta[1,] = mu0
Sigma[,,1] = S0
### Define functions for certain variables in the full conditional distribution
An <- function (Sigma) {
  solve(Sigma0) + n * solve(Sigma)
}
bn <- function (Sigma) {
  solve(Sigma0) %*% mu0 + solve(Sigma) %*% Y.sum
}
mu_n <- function (Sigma) {
  solve(An(Sigma)) %*% bn(Sigma)
}
```

```r
Sigma_n <- function (Sigma) {
  solve(An(Sigma))
}
S_theta <- function (theta) {
  data.matrix(t(Y-theta)) %*% data.matrix(Y-theta)
}
### Run the Gibbs sampler for the intended number of iterations
for (s in 2:n.iter) {
  theta[s,] = rmvnorm(1, mu_n(Sigma[,,s-1]), Sigma_n(Sigma[,,s-1]))
  Sigma[,,s] = solve(rWishart(1, nu.n, solve(Sigma0 + S_theta(theta[s,])))[,,1])
}
### Repeat for orange crab
Y = data_orangecrab
set.seed(12)
mu0 = colMeans(Y)
S0 = cov(Y)
Y.sum = colSums(Y)
n = nrow(Y)
theta_orange = array(dim = c(n.iter, 2))
Sigma_orange = array(dim = c(2, 2, n.iter))
theta_orange[1,] = mu0
Sigma_orange[,,1] = S0
for (s in 2:n.iter) {
  theta_orange[s,] = rmvnorm(1, mu_n(Sigma_orange[,,s-1]), Sigma_n(Sigma_orange[,,s-1]))
  Sigma_orange[,,s] = solve(rWishart(1, nu.n,
                                     solve(Sigma0 + S_theta(theta_orange[s,])))[,,1])
}
```
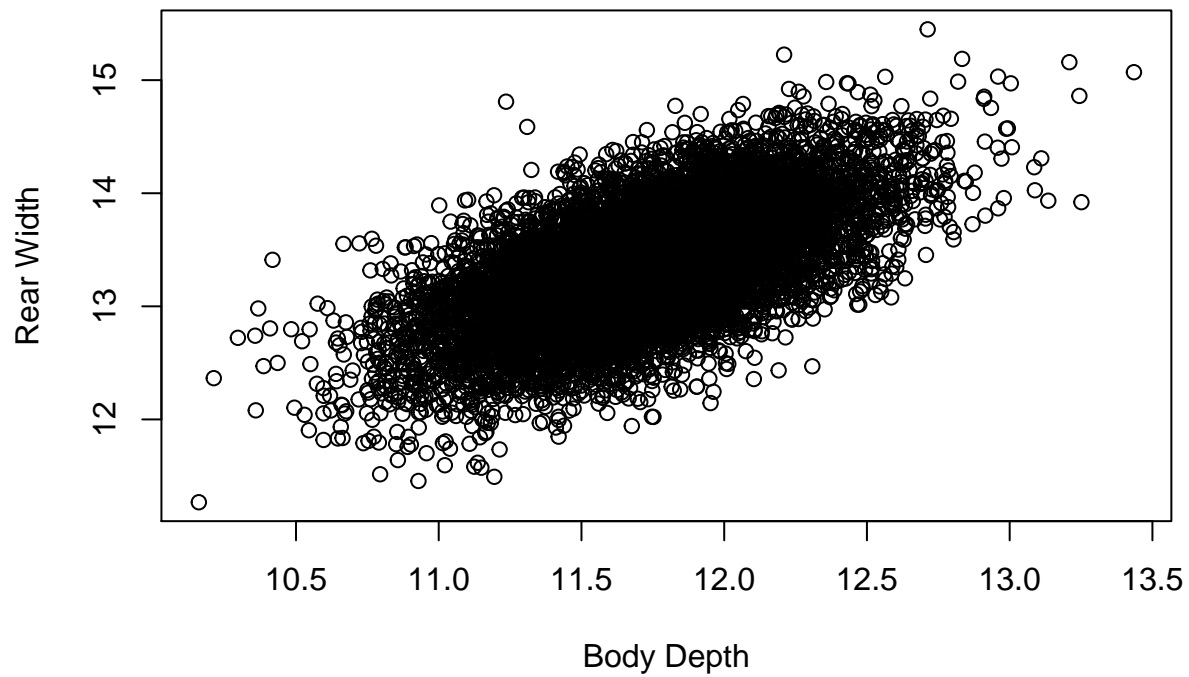
   b.

```r
### Plot blue crab theta values
plot(theta[,1], theta[,2], xlab = "Body Depth", ylab = "Rear Width",
     main = "Blue Crab Population Mean Samples")
```
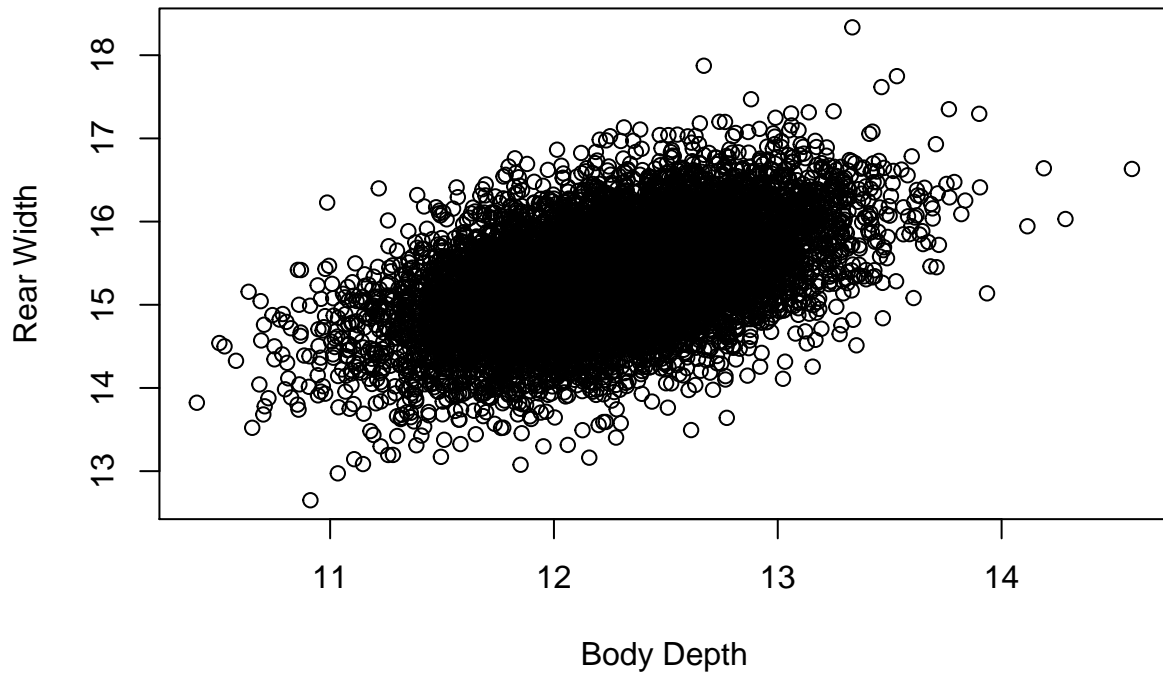
Blue Crab Population Mean Samples


```r
### Plot orange crab theta values
plot(theta_orange[,1], theta_orange[,2], xlab = "Body Depth", ylab = "Rear Width",
     main = "Orange Crab Population Mean Samples")
```

Orange Crab Population Mean Samples

There are some small size differences between the two groups. On average, orange crabs have more body depth and rear width than blue crabs do. As shown in the graphs above, orange crabs have a body depth of about 12.5 and a rear width of about 15.5, while blue crabs have a body depth of about 11.75 and a rear width of about 13.5. Both groups show a moderately strong positive correlation between body depth and rear width.

c.

```
### Plot posterior density of correlation coefficients for blue crabs
rho_blue = sapply(1:n.iter, function (i) {
  S = Sigma[,,i]
  rho = S[1,2] / sqrt(S[1,1]*S[2,2])
  return (rho)
})
plot(density(rho_blue), xlab = "Correlation Coefficient", ylab = "",
     main = "Posterior Density of Blue Crab Correlation Coefficients")
```

Posterior Density of Blue Crab Correlation Coefficients
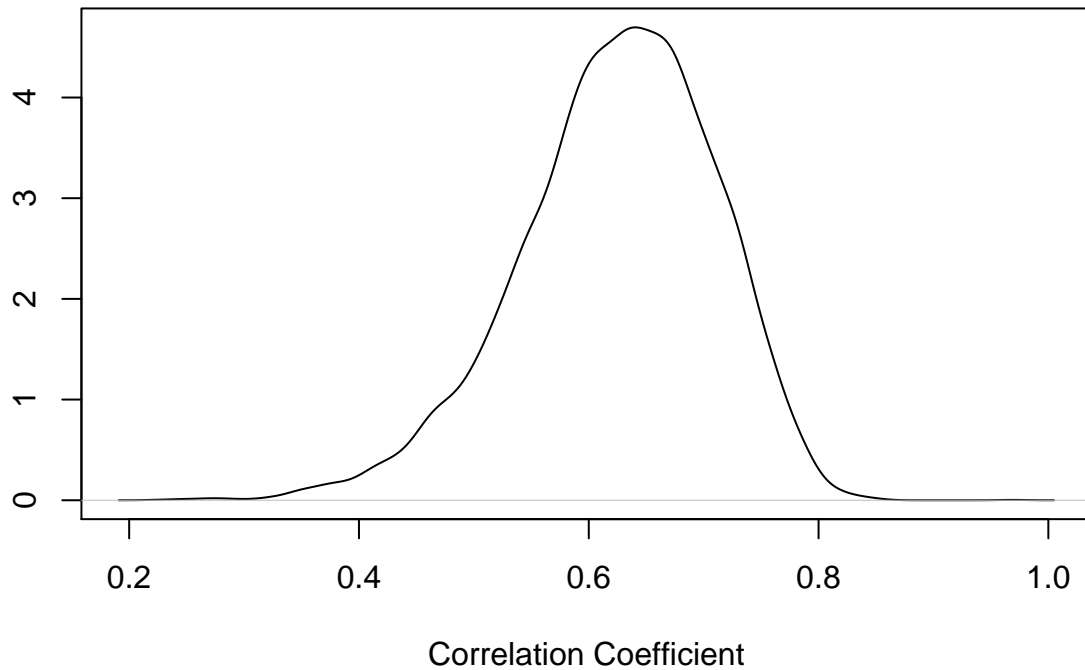

```
### Plot posterior density of correlation coefficients for orange crabs
rho_orange = sapply(1:n.iter, function (i) {
  S = Sigma_orange[,,i]
  rho = S[1,2] / sqrt(S[1,1]*S[2,2])
  return (rho)
})
plot(density(rho_orange), xlab = "Correlation Coefficient", ylab = "",
    main = "Posterior Density of Orange Crab Correlation Coefficients")
```

Posterior Density of Orange Crab Correlation Coefficients
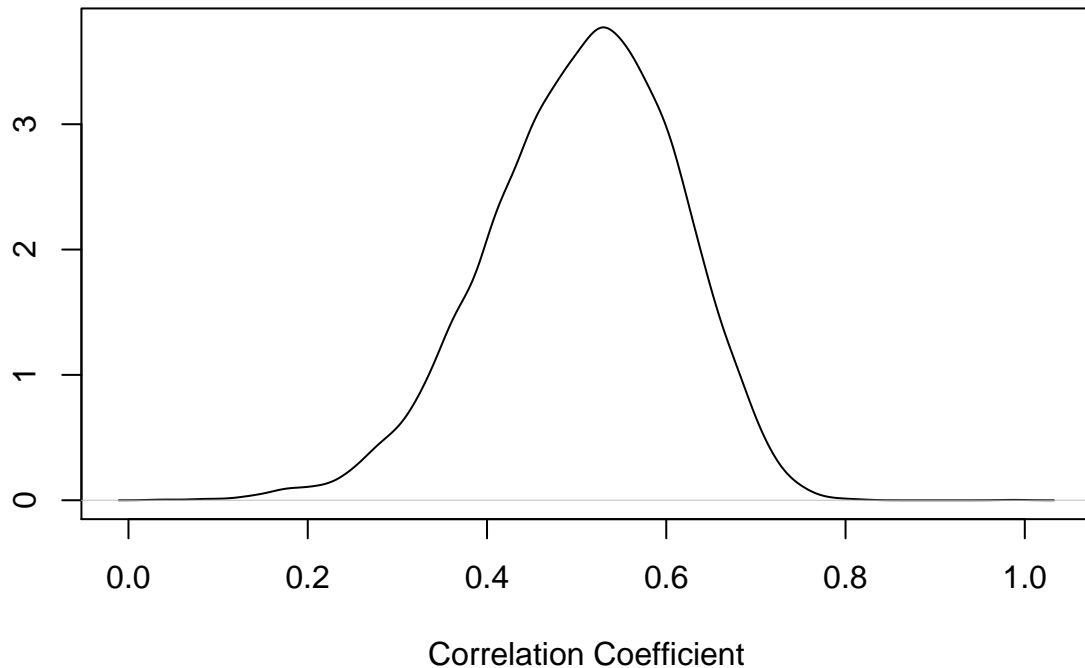

```
### Finding the probability that rho_orange > rho_blue
probability <- sum((rho_blue < rho_orange))/n.iter
probability
```

```
## [1] 0.1857
```

These results suggest that blue crabs exhibit a stronger positive correlation between body depth and rear width than orange crabs do. This is because the probability that the correlation coefficient for the blue crabs is less than the correlation coefficient for the orange crabs is estimated to be 18.57%, so the probability that the correlation coefficient for the blue crabs is greater than or equal to the correlation coefficient for the orange crabs is estimated to be 81.43%. From these probabilities, it is clear that it is more likely for blue crabs to exhibit a stronger positive correlation between body depth and rear width than orange crabs do than for orange crabs to exhibit a stronger positive correlation between body depth and rear width than blue crabs do.