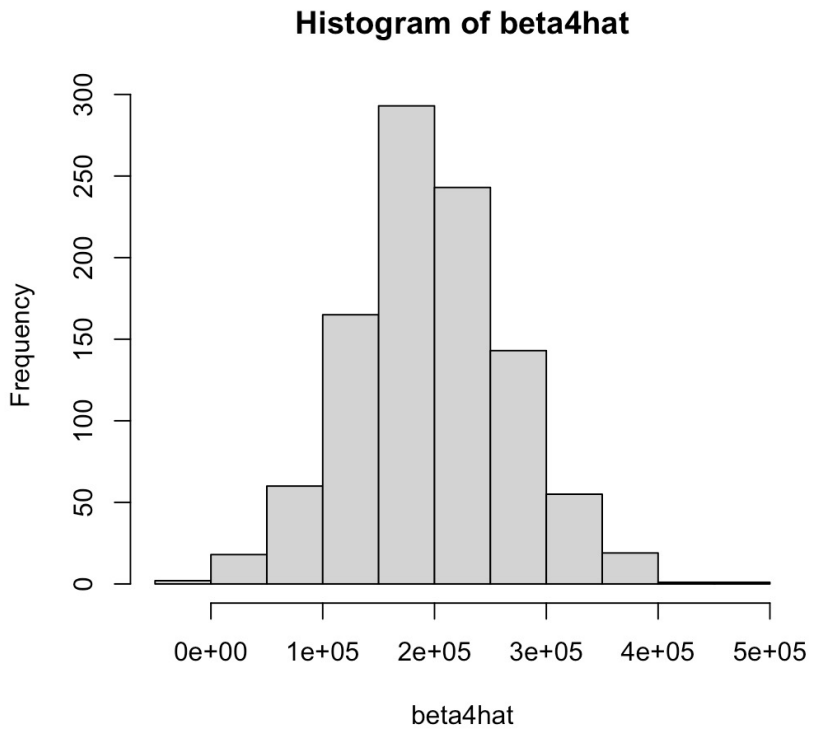
Question 1

a) i)

```
# Question 1
setwd("~/Desktop/STAT 371/a1")
name=read.table("A1_variates.txt",header = TRUE)
attach(name)
set.seed(20884580)

Beta <- c(-250000, 50, -200, 12000, 200000)
X=as.matrix(cbind(rep(1,24),name))
beta = array(0,c(5,1000)) #creates an array to store sets of estimators
for (i in 1:1000){
   y<-X%*%Beta+rnorm(24,0,15000) #X and Beta as defined in 2a) of A1
   A1sim.lm<-lm(y~size+age+employees+col)
   beta[,i]<-coef(A1sim.lm) #yields the estimates for each iteration
}


# a)
# i)
beta4hat <- beta[5,]
hist(beta4hat)
```



ii) The sample mean is 196333.9.

```
# ii)
mean(beta[5,])
```

iii) The sample standard deviation is 70934.06.

```
# iii)
sd(beta[5,])
```

iv) By below R code and question 2 from assignment 1, we know that $\widehat{\beta_4}$ ~ N(200000, 5091094863). From part i), we observe that there is normal distribution. To be specific, it is roughly symmetric and there are a lot of data around the center and less data around the two tails. Thus, the empirical result from part i) is consistent with the distribution of $\widehat{\beta_4}$ derived in class. From part ii), we find that the sample mean is 196333.9, which is close to 200000. Thus, the empirical result from part ii) is consistent with the distribution of $\widehat{\beta_4}$ derived in class. From part iii), we find that the sample standard deviation is 70934.06. By writing R code below, we get 71351.91. These two values are close and so the empirical result from part iii) is consistent with the distribution of $\widehat{\beta_4}$ derived in class.

```
# iv)
model.matrix(A1sim.lm)
XtXinv = solve(t(X)%*%X) #yields the (XtX)-1 matrix
varb4 <- XtXinv[5,5]*15000^2
varb4
sdb4 <- sqrt(varb4)
sdb4
```