Assignment 2 Question 4

a).

```
# generate two predictor variables
x1 <- rnorm(200, mean=0, sd=2)
x2 <- rnorm(200, mean=0, sd=2)

# generate epsilon
eps <- rnorm(200, mean=0, sd=1)

# generate response vector y
y = 1 + 2*x1 + 5*x2 + eps

# estimate the least squares regression line
reg <- lm(y ~ x1+x2)
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9701 -0.7009 -0.1208  0.5994  2.4140
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.99746    0.06758   14.76   <2e-16 ***
## x1           2.00097    0.03538   56.56   <2e-16 ***
## x2           5.02308    0.03421  146.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9542 on 197 degrees of freedom
## Multiple R-squared:  0.9931, Adjusted R-squared:  0.993
## F-statistic: 1.413e+04 on 2 and 197 DF,  p-value: < 2.2e-16
```

the estimated line is:

```
paste("y=",
      reg$coefficients[1],"+",
      reg$coefficients[2],"x1", "+",
      reg$coefficients[3],"x2")
```

```
## [1] "y= 0.997461285142567 + 2.00097125767919 x1 + 5.02308295195223 x2"
```

b).

```
# estimate the variance for each of the three regression coefficients
reg_sum <-summary(reg)
(reg_sum$coefficients[1,2])^2
```

```
## [1] 0.004567332
```

```
(reg_sum$coefficients[2,2])^2
```

```
## [1] 0.001251563
```

```
(reg_sum$coefficients[3,2])^2
```

```
## [1] 0.001170202
```

```
# theoretical values for the variance of these predictors
X <- cbind(rep(1, 200), x1, x2)
diag(solve(t(X) %*% X))
```

```
##                      x1          x2
## 0.005015789 0.001374452 0.001285102
```

the estimated variance of the three regression coefficients are close to theoretical values for the variance of these predictors.

c).

```
paste("p-value is",
      reg_sum$coefficients[2,4])
```

```
## [1] "p-value is 9.33781479299284e-124"
```

p-value is less than significant level 0.05, so we reject null hypothesis. we can conclude that the coefficient for x1 is not zero.

d).

```
# repeat part a. and c. 1000 times
count <- 0
for(i in 1:1000){
  x1 <- rnorm(200, mean=0, sd=2)
  x2 <- rnorm(200, mean=0, sd=2)
  eps <- rnorm(200, mean=0, sd=1)
  y = 1 + 2*x1 + 5*x2 + eps
  reg <- lm(y ~ x1+x2)
  reg_sum <- summary(reg)
  pval <- reg_sum$coefficients[2,4]
  if (pval < 0.05) count <- count + 1
}
print(count)
```

```
## [1] 1000
```

repeat part a and c 1000 times, we reject null hypothesis 1000 times.