Question4

```
set.seed(12345678)
data1 <- rchisq(100 , 5, ncp = 0)
result.mean <- mean(data1)
data2 <- rchisq(10000 , 5, ncp = 0)
result.mean2 <- mean(data2)
print(result.mean)
```

```
## [1] 5.201113
```

```
print(result.mean2)
```

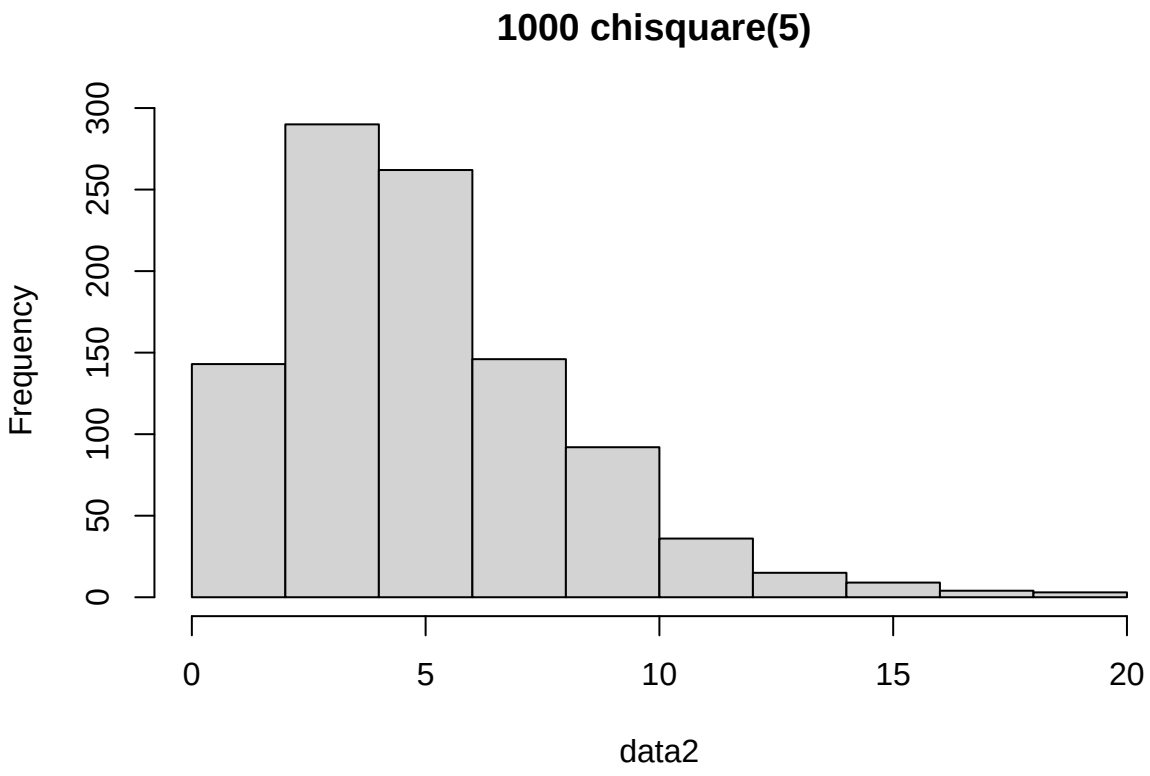
```
## [1] 5.007759
```

The mean of rchisq(10000 , 5) is closer to population mean, which can be explained by WLLN.
The following is the histogram of rchisq(1000,5)

```
set.seed(12345678)
data2 <- rchisq(1000 , 5, ncp = 0)
var(data2)
```

```
## [1] 9.399373
```

```
hist(data2,main="1000 chisquare(5) ")
```



The following is the histogram of the 1000 means

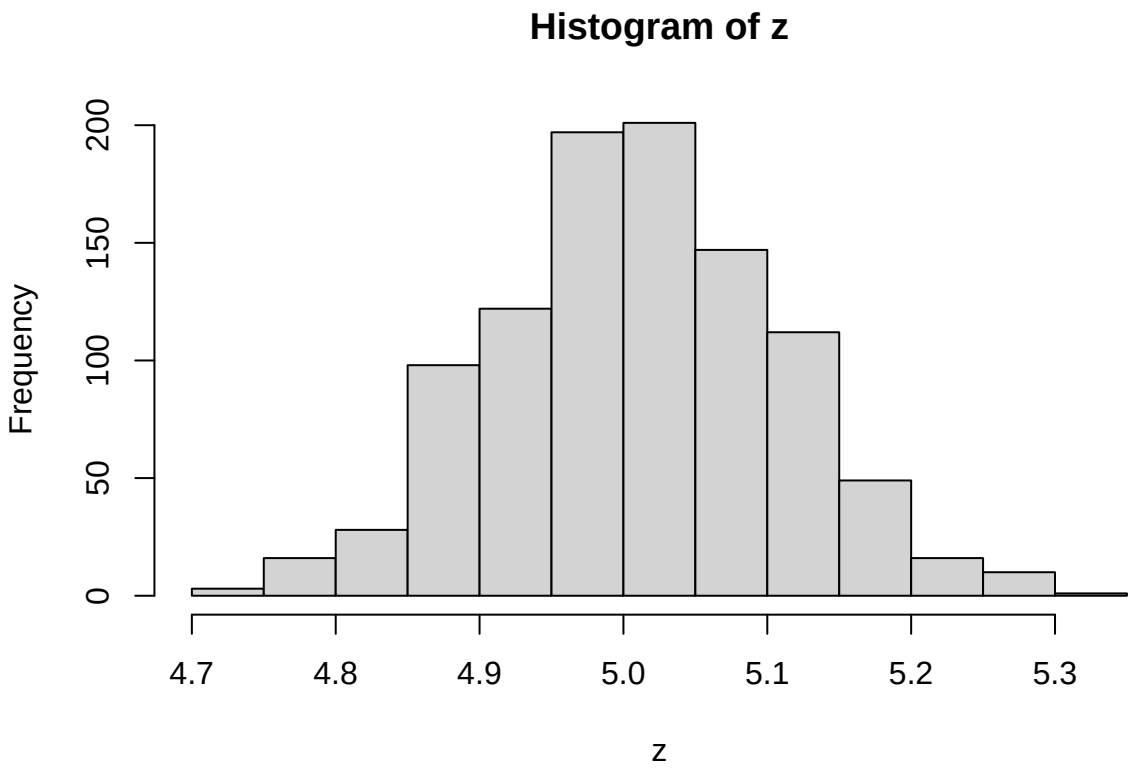
```
set.seed(12345678)
z <- c(1:1000)
for( i in z){
  dataTemp <- rchisq(1000 , 5, ncp = 0)
  z[i] <-mean(dataTemp)
}
var(z)
```

```
## [1] 0.009851764
```

```
mean(z)
```

```
## [1] 5.008624
```

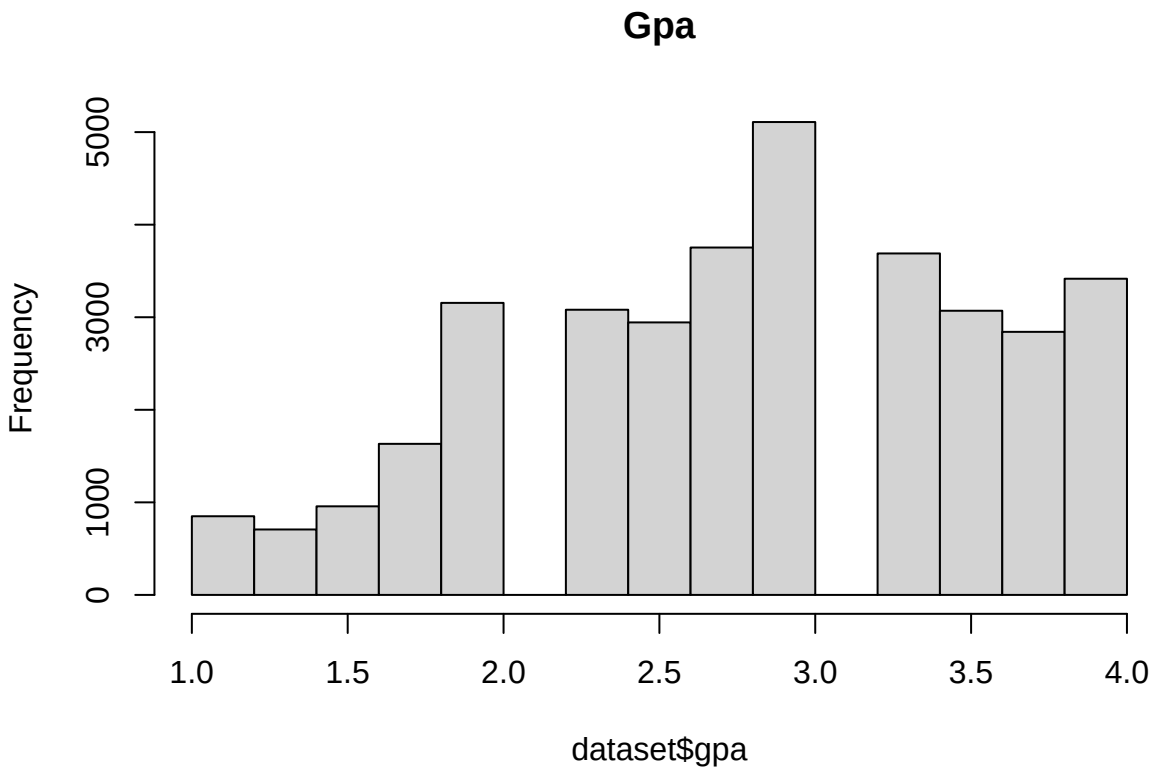
```
hist(z)
```



Z is a normal distribution. The variance and mean are identical to what CLT predict.(mean(z) = mean(data2), var(z) = var(data2)/1000)

Question 5

```
library(readxl)
dataset <- read_excel("student_sample.xlsx")
hist(dataset$gpa, main="Gpa")
```



```r
reg1 <-lm(gpa~black+Asian+deliquent, data = dataset)
summary(reg1)
```

```
##
## Call:
## lm(formula = gpa ~ black + Asian + deliquent, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23958 -0.51645  0.05806  0.58024  2.00906
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.003369   0.006800  441.64   <2e-16 ***
## black       -0.275240   0.011040  -24.93   <2e-16 ***
## Asian        0.236210   0.015512   15.23   <2e-16 ***
## deliquent   -0.122865   0.003829  -32.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7513 on 35201 degrees of freedom
## Multiple R-squared:  0.05052,    Adjusted R-squared:  0.05044
## F-statistic: 624.3 on 3 and 35201 DF,  p-value: < 2.2e-16
```

In this dataset, the black come with worse gpa. I think it is because the black in general dosen't have the resource to improve their grade, such as tutor, compared to the other race.
In this dataset, Asian come with better gpa. I think it is because Asian parents in general are more serious about the grade of their children.
Beside race, delinquency comes with poor gpa in this dataset. I think the worse grade is because they are busy doing else.