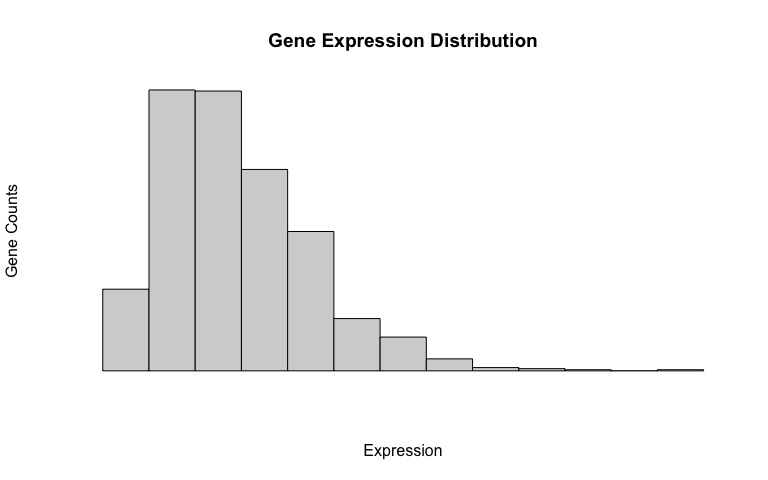
# Biostatistics Homework 4

By umi (张鹿鸣12112618)

## 1 Gene Expression Distribution

### 1.1)

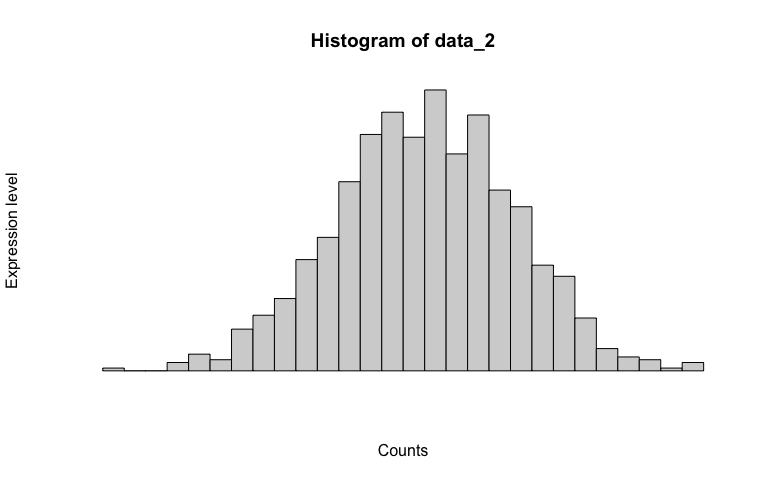
data <- rgamma(1000, 3,1)  
hist(data,  
 main = 'Gene Expression Distribution',  
 xlab = 'Expression',  
 ylab = 'Gene Counts',  
 xaxt = 'n',  
 yaxt = 'n')



I think the gene expression pattern will follow a Gamma distribution, which in other words is a distribution skewed to the right. Because many genes are not significantly expressed, thus most of the genes will have a rather low expression level. But some genes are expressed in a very significant amount, so a few counts will lie in the high expression area. Thus I think a Gamma distribution will fit best, with most of the values lie on the left, and a few values lie to the right.

### 1.2)

data\_2 <- rnorm(1000, 5, 0.4)  
hist(data\_2,  
 breaks = 20,  
 xaxt = 'n', xlab = 'Counts',  
 yaxt = 'n', ylab = 'Expression level')



I think the *Nanog* gene expression across the embryonic stem cell population should follow a normal distribution. Because they are all from the same embryonic stem cell population, their expression of a particular gene should not vary from a certain value vary dramatically, although there might be fluctuations.

### 1.3)