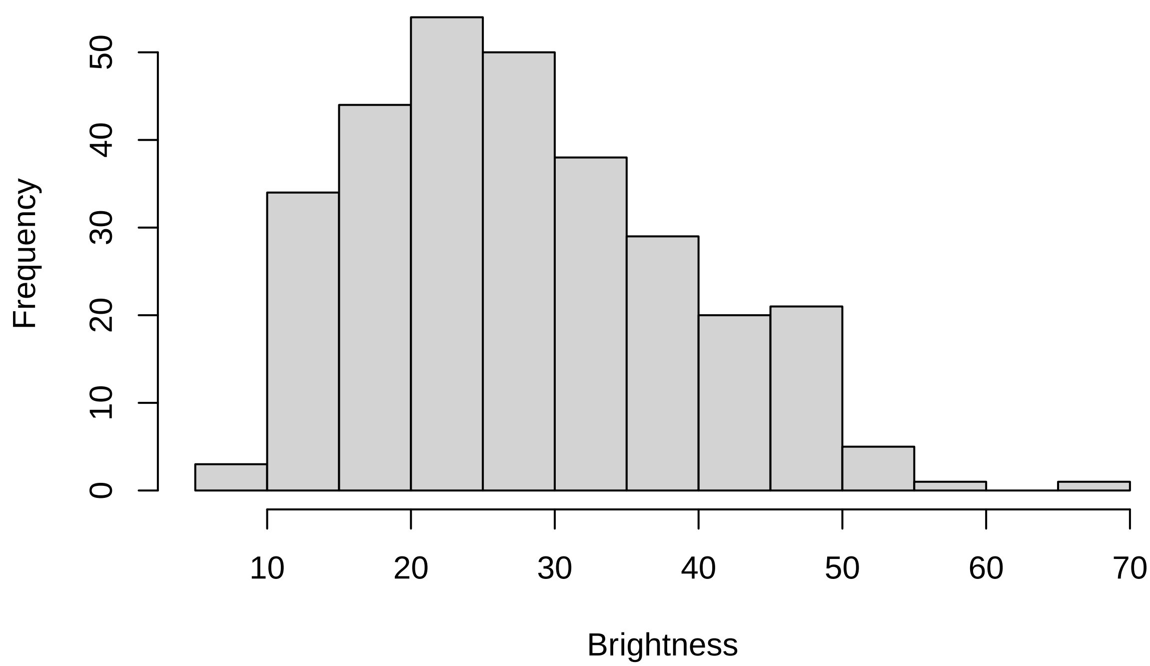
A5Q1

a)

i)

```r
digit <- read.csv('digitData.csv',header=T)
brightness <- digit$Brightness
hist(brightness, breaks='FD', xlab='Brightness', main='')
```



ii)

```r
c(mean(brightness), skew(brightness))
```

```
## [1] 27.9970493  0.5361338
```

iii)

```
createSkewFunction <- function(y.pop) {
  skew2nd <- function(alpha) {
    skew( powerfun(x=y.pop, alpha) )
  }
}

brightness.sk = createSkewFunction(brightness)
alpha.star = uniroot(brightness.sk, interval=c(-2,2))$root
alpha.star
```

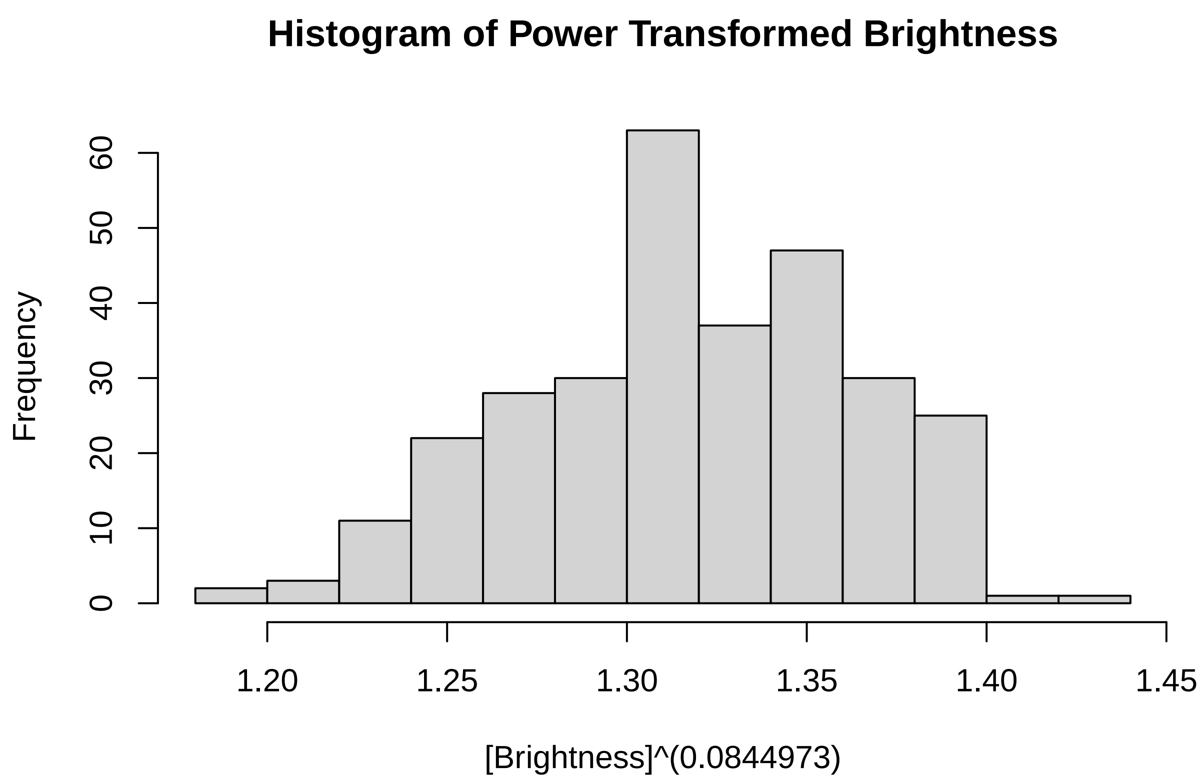
```
## [1] 0.0844973
```

iv)

```
skew(powerfun(brightness, alpha.star))
```

```
## [1] -1.834623e-08
```

```
hist(powerfun(brightness, alpha.star), xlab="[Brightness]^(0.0844973)",
     main="Histogram of Power Transformed Brightness", breaks='FD')
```



v)

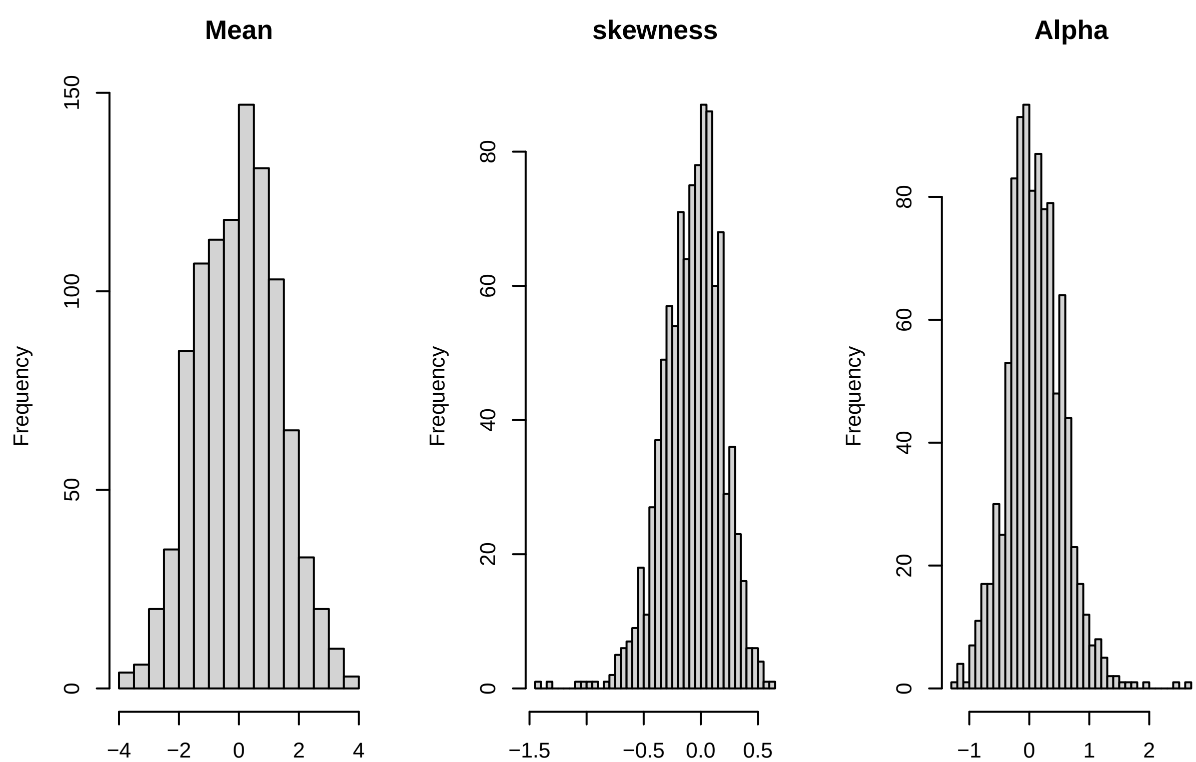
```
attr3 <- function(z){
  alpha.star = uniroot(createSkewFunction(z), interval=c(-8,8))$root
  c(mean(z), skew(z), alpha.star)
}
attr3(brightness)
```

```
## [1] 27.99704932  0.53613378  0.08449728
```

b)

```
M= 1000
n = 50
samples <- sapply(1:M, function(m){sample(brightness, n, replace=FALSE)})
sample.attr <- apply(samples, 2, attr3)
pop.attr <- attr3(brightness)

par(mfrow=c(1,3))
hist(sample.attr[1,]-pop.attr[1], breaks='FD', main='Mean', xlab='')
hist(sample.attr[2,]-pop.attr[2], breaks='FD', main='skewness', xlab='')
hist(sample.attr[3,]-pop.attr[3], breaks='FD', main='Alpha', xlab='')
```



c)

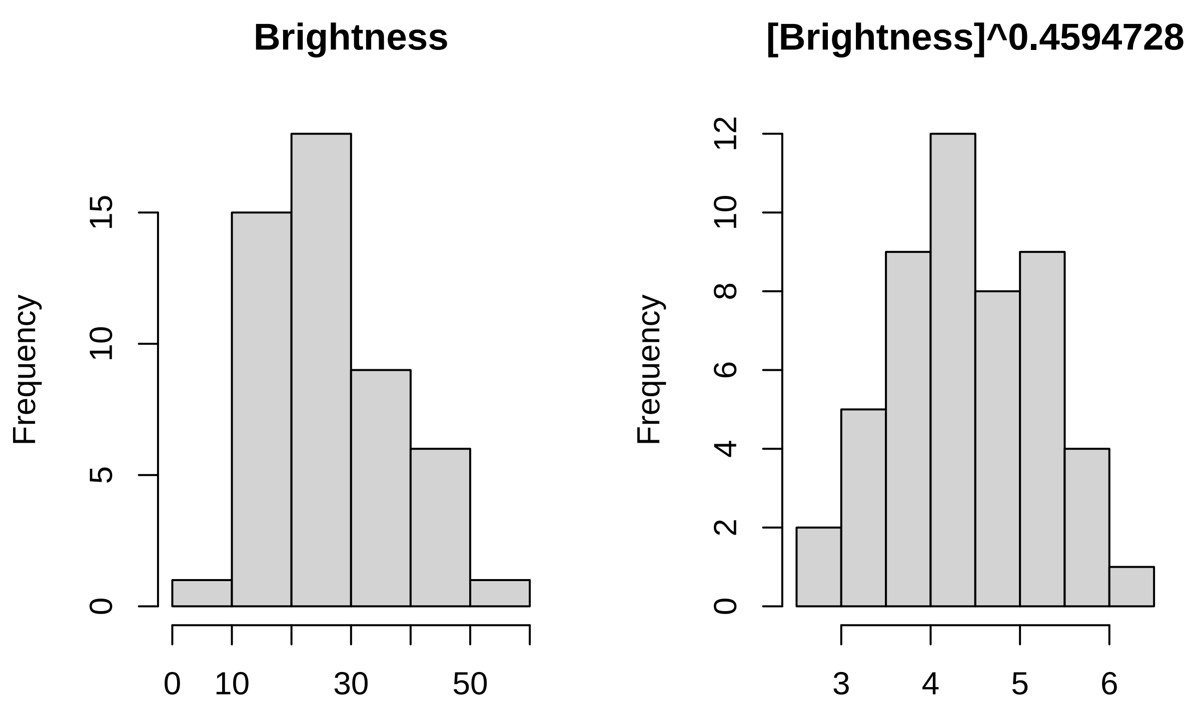
i)

```
digitSample <- c(294,133,95,265,154,1,289,232,121,99,129,83,30,56,249,134,46,68,
                 165,279,105,91,248,285,238,45,194,34,44,5,173,87,18,299,167,64,
                 42,266,281,210,27,207,271,181,6,212,176,51,28,243)
brightSam <- brightness[digitSample]
bright.attr <- attr3(brightSam)
bright.attr
```

```
## [1] 26.6065306  0.3144501  0.4594728
```

ii)

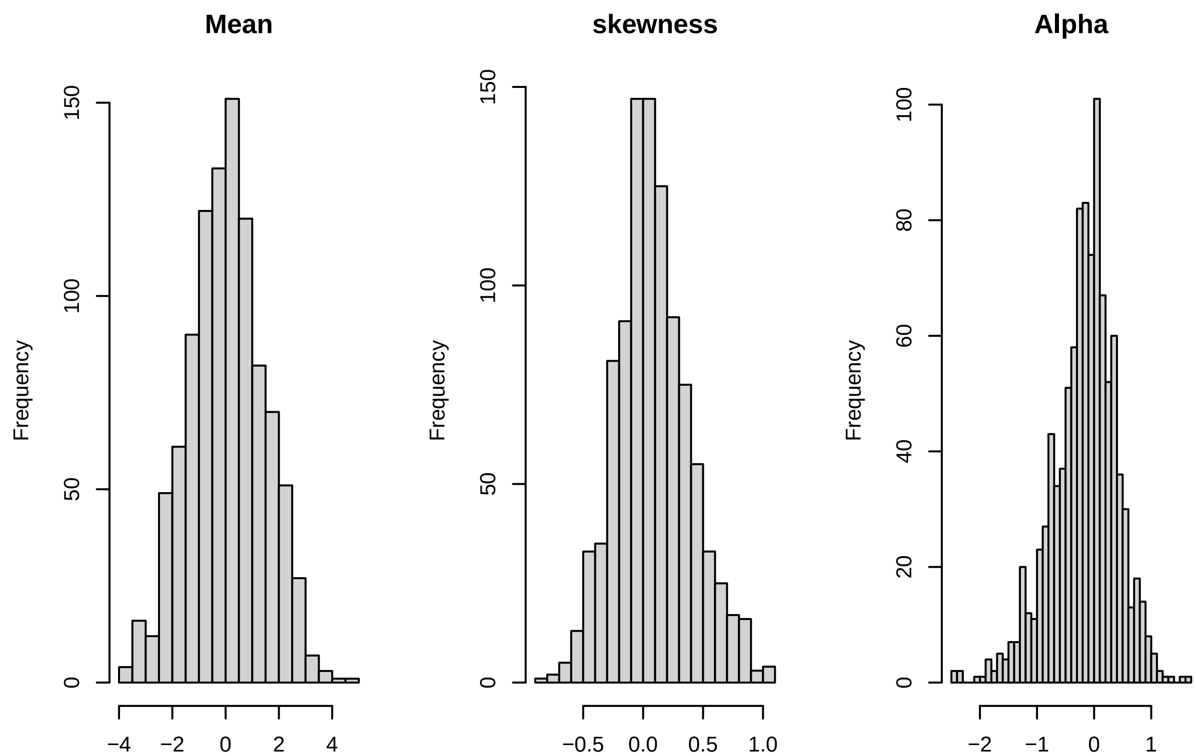
```
par(mfrow=c(1,2))
hist(brightSam, breaks='FD', xlab='', main='Brightness')
hist(powerfun(brightSam, bright.attr[3]), breaks='FD', xlab='',
     main='[Brightness]^0.4594728')
```



iii)

```
B = 1000
s.star <- sapply(1:B, function(m){sample(brightSam, replace=TRUE)})
boot.attr <- apply(s.star, 2, attr3)

par(mfrow=c(1,3))
hist(boot.attr[1,]-bright.attr[1], breaks='FD', main='Mean', xlab='')
hist(boot.attr[2,]-bright.attr[2], breaks='FD', main='skewness', xlab='')
hist(boot.attr[3,]-bright.attr[3], breaks='FD', main='Alpha', xlab='')
```



iv)

```
s <- matrix(0, nrow= 3, ncol= 3)
rownames(s) <- c('Mean', 'Skewness', 'Alpha')
colnames(s) <- c('std error', 'Lower', 'Upper')

s[, 1] <- apply(boot.attr, 1, sd)
s[, 2:3] <- t(apply(boot.attr, 1, quantile, c(0.025, 0.975)))
s
```

```
##           std error      Lower     Upper
```

```
## Mean     1.4161685 23.8475485 29.407819
## Skewness 0.3089603 -0.1607520  1.085058
## Alpha    0.5688230 -0.9860346  1.291386
```

d)

```r
M = 100
B = 1000
n = 50

samples <- sapply(1:M, function(m){sample(brightness, n, replace=FALSE)})
sample.conf <- apply(samples, 2, function(sm){
  s.star <- sapply(1:B, function(m){sample(sm, replace=TRUE)})
  boot.attr <- apply(s.star, 2, attr3)
  apply(boot.attr, 1, quantile, c(0.025, 0.975))
})
```

```r
p <- matrix(0, nrow=3, ncol=2)
rownames(p) <- c('Mean', 'Skewness', 'Alpha')
colnames(p) <- c('Converge Estimate', 'Std Error')

p[1,1] <- mean((pop.attr[1]>=sample.conf[1,]) & (pop.attr[1]<=sample.conf[2,]))
p[2,1] <- mean((pop.attr[2]>=sample.conf[3,]) & (pop.attr[2]<=sample.conf[4,]))
p[3,1] <- mean((pop.attr[3]>=sample.conf[5,]) & (pop.attr[3]<=sample.conf[6,]))
p[,2] <- sqrt(p[,1]*(1-p[,1])/M)
p
```

```
##          Converge Estimate  Std Error
## Mean                  0.97 0.01705872
## Skewness              0.98 0.01400000
## Alpha                 0.97 0.01705872
```

Conclusion: Since the percentages of the bootstrap interval coverage of three attributes are all over 95%, three attributes all over estimate the proportion instead of expected 95%.