hw5

Kyu Park

2021 2 14

5.

a)

```r
load("~/STAT23400/Wealth.Rdata")
overnums = nrow(subset(Wealth, Y2017 > 5000000)) ##counts above 5 million
totalnums = nrow(subset(Wealth, Y2017 <= max(Wealth$Y2017, na.rm = TRUE))) ##total population count
proportion = overnums/totalnums
proportion
```

```
## [1] 0.005100448
```

The estimate proportion is 0.005100448. 95% confidence interval could be calculated as

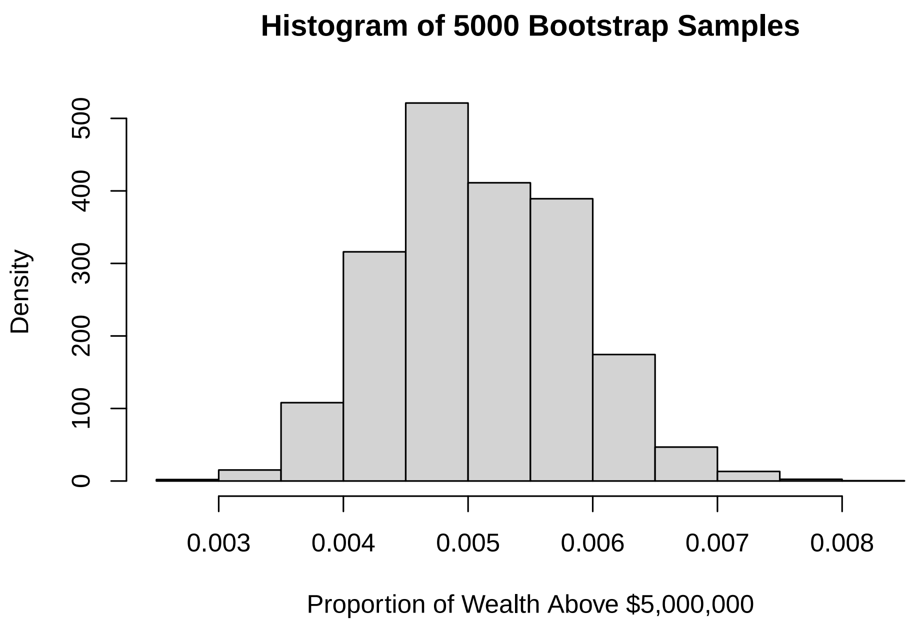
$$\bar{x} \pm z\sqrt{\frac{p(1-p)}{n}}$$

, where $\bar{x} = 0.005100448$ and $p = 0.005100448$, $z = 1.96$, $n = 9607$. This yields CI of (0.00367, 0.00652).

b) Assume that the sample size of 9607 is sufficiently large enough to make an estimate, and that the distribution will be approximately normal, and therefore there will be no clear skewedness.

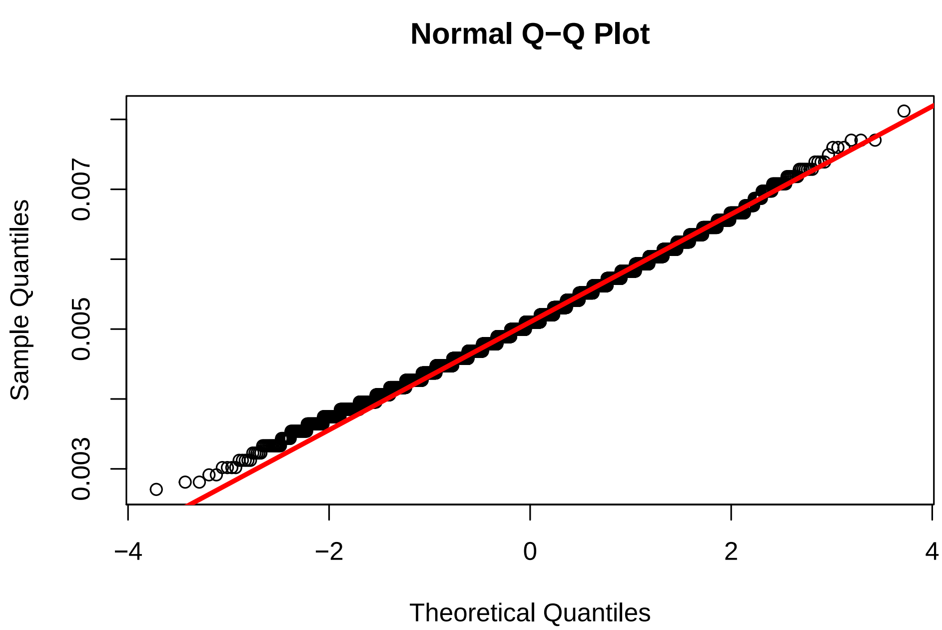
c)

```r
##remove all the NULLs
wealth2017 = t(subset(Wealth, Y2017 <= max(Wealth$Y2017, na.rm = TRUE), select = "Y2017"))
##5000 samples of size 9607
samp = t(replicate(5000, sample(wealth2017, 9607, replace = TRUE)))
##vector of the proportion of the sample above $5,000,000
bootprop = rowSums(samp > 5000000)/9607
hist(bootprop, freq = FALSE, breaks = 10,
     xlab = "Proportion of Wealth Above $5,000,000",
     main = "Histogram of 5000 Bootstrap Samples")
```



By the central limit theorem, as the bootstrap sample sizes increase, the distribution must look normal. The distribution of the graph above looks normal with sample size of 5000, as expected. The normality of the bootstrap samples could be shown by the quantile plot.

```
qqnorm(bootprop)
qqline(bootprop, col = "red", lwd = 3)
```



d)

```r
quantile(bootprop, c(0.025, 0.975))
```

```
##        2.5%       97.5%
## 0.003747268 0.006557718
```

The CI interval of my bootstrap sample in 95% confidence level is shown above, which is about the same with a).