**HW 1 - Avinash Ramu**

Question 2

Part I

proportions <- c(.4777, .4875, .4859, .4754, .4874, .4864, .4813, .4787, .4895, .4797, .4876, .4859, .4857, .4907, .5010, .4903, .4860, .4911, .4871, .4725, .4822, .4870, .4823, .4973)

The observed standard deviation is given by:

sd(proportions) = 0.006409724

Expected theoretical standard deviation = sqrt(p \* (1-p) / n)

Here p = mean(proportions) = 0.485675, n = 3900

=> expected SD = 0.0080

Part II

From theory:

s^2 \* (n - 1) / sigma^2 follows a chi-squared distribution with n-1 degrees of freedom. s is the sample SD and sigma is the expected SD.

The 95% confidence interval can be calculated as sqrt(0.008 \* 0.008 \* qchisq(0.05, 23) / 23), sqrt(0.008 \* 0.008 \* qchisq(0.05, 23) / 23)

The 95% confidence interval comes out to (0.006035259, 0.009892641)

The observed SD lies within this interval and hence the observed difference is not statistically significant.

Question III

for (i in 1:1000) { x <- c(x, sum(rnorm(20))) }

hist(x, freq = F)

curve(dnorm(x, mean = mean(x), sd = sd(x)), add = T)

The density curve is more spread out than the histogram, this is because the probability mass is spread across the tails from neg infinity to infinity. The histogram has a higher peak which might go down with an increasing sample size.