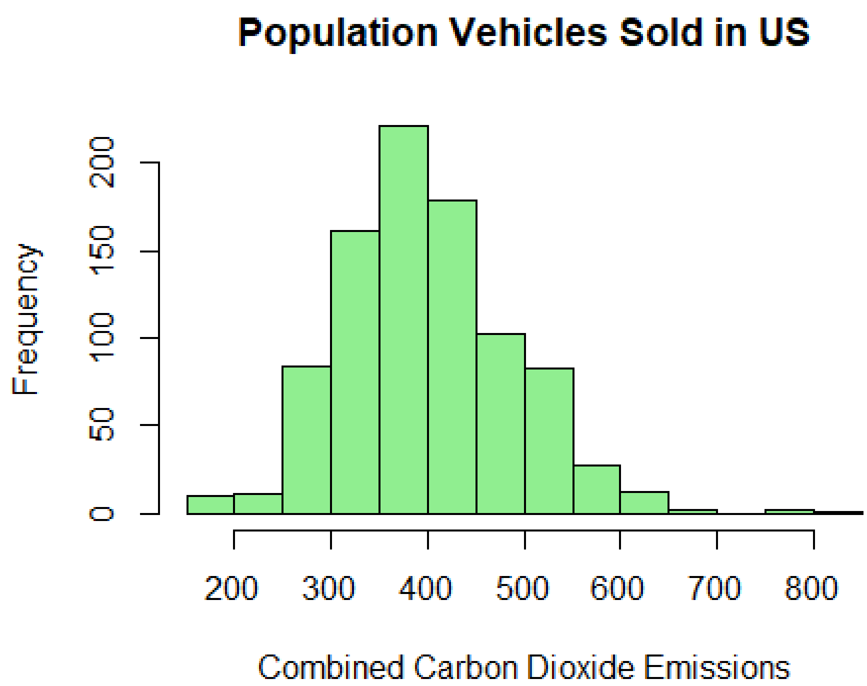
ST 314 - Data Analysis 4

Ivan Timothy Halim

February 4, 2019

Part 1

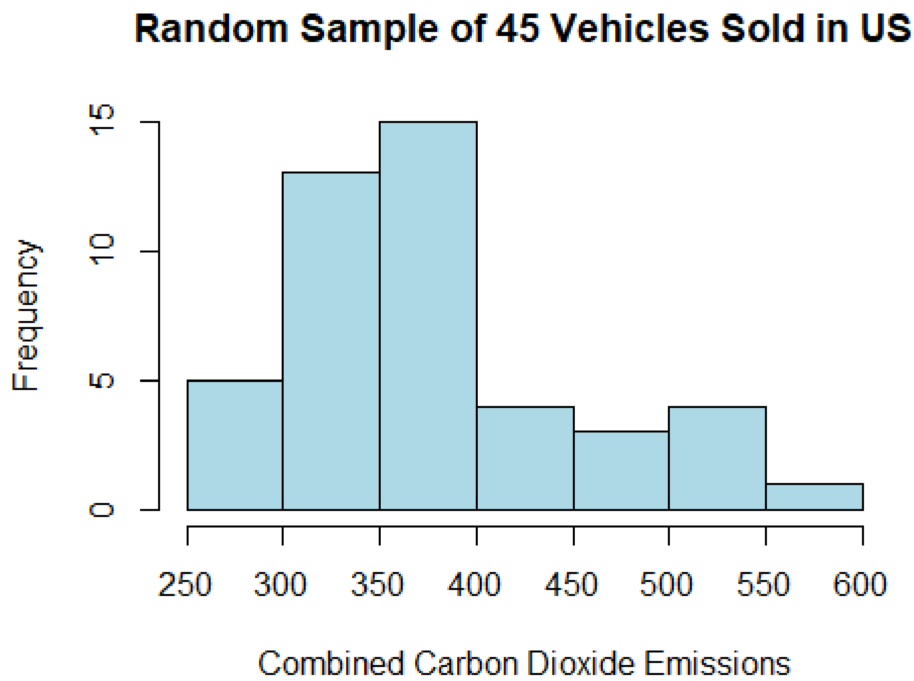
a.



$N = 896$
$\mu = 399.87$
$\sigma = 89.83$

The population data seems to be unimodal and normally distributed, ranging from about 150 to 850 with many observation hovering between 350 and 400. There is a small number of extreme observations ranging from 750 to 850.

b.



$\overline{x} = 393.47$
$s = 84.91$

Compared to the population data, the sampled data seems to be slightly positively skewed.

c. $90\%\ CI = \overline{x} \pm 1.645 \times \left(\frac{\sigma}{\sqrt{n}}\right) = 393.47 \pm 22.03 = (371.44, 415.49)$

The confidence interval includes the true population mean.

d. For 90% Confidence Interval, about 10% of students will not capture the true population mean, which is about 26 students.

Part 2

a. We will most likely fail to reject the null hypothesis, because the true population mean is within 0.5 standard error from our point estimate, which is very close.

b. $z = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{393.47-399.87}{\frac{89.83}{\sqrt{45}}} = -0.48$

$p\text{-}value = 2 \times 0.3156\ (by\ z\text{-}table) = 0.6312$

c. There is insufficient evidence to conclude that the true population differs from 399.87.

Fail to reject the null hypothesis at a significance level of 0.10 (z stat = -0.48, p-value = 0.6312).

The sample estimates the true average combined carbon dioxide emission to be 393.47 with a 90% confidence interval of 371.44 to 415.49.

The sample is fairly accurate.

d. If the confidence interval does not contain the true population mean, then the z-score will be above 1.645, which means that the p-value will be less than the significance level 0.10. Therefore, the sample will also reject the null hypothesis.

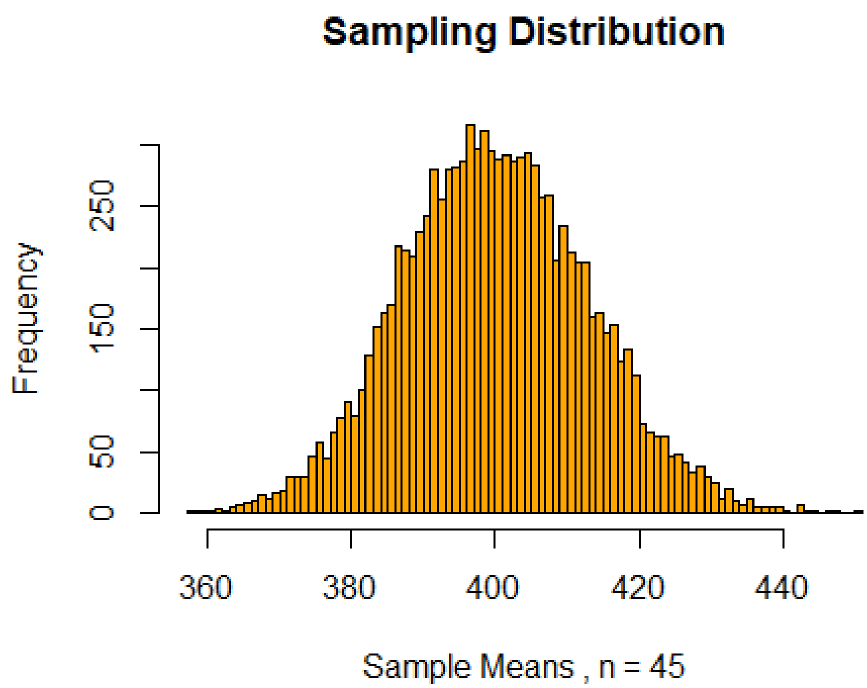
Part 3

a. According to the Central Limit Theorem, the distribution of the sample mean should be a normal distribution centered at μ with a standard error of $\sigma/\sqrt{n}$

$$\mu_{\overline{x}} = \mu = 399.87$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{89.83}{\sqrt{45}} = 13.39$$

b.

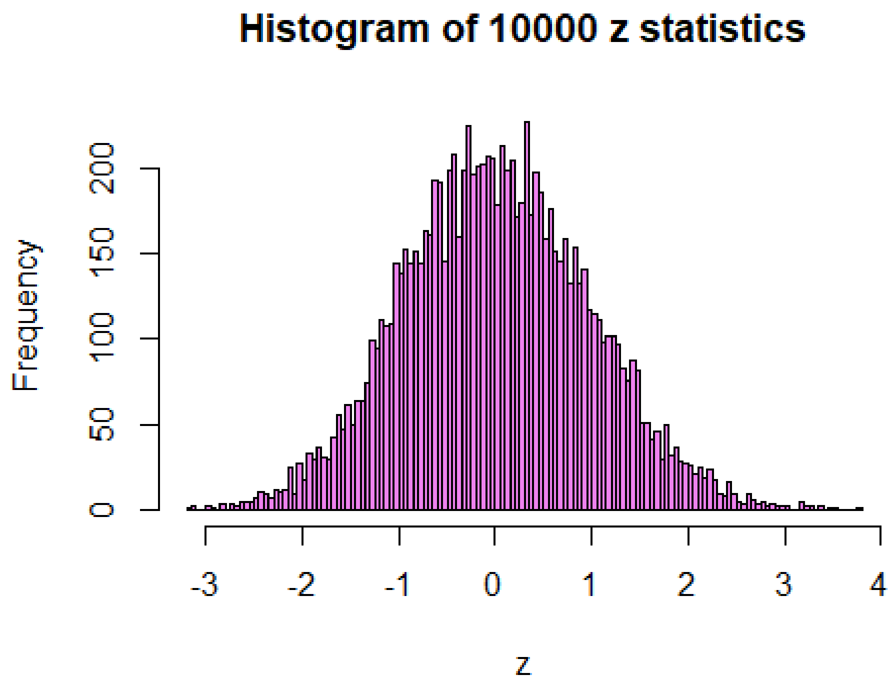


$$\mu_{\overline{x}} = 399.87$$

$$\sigma_{\overline{x}} = 13.19$$

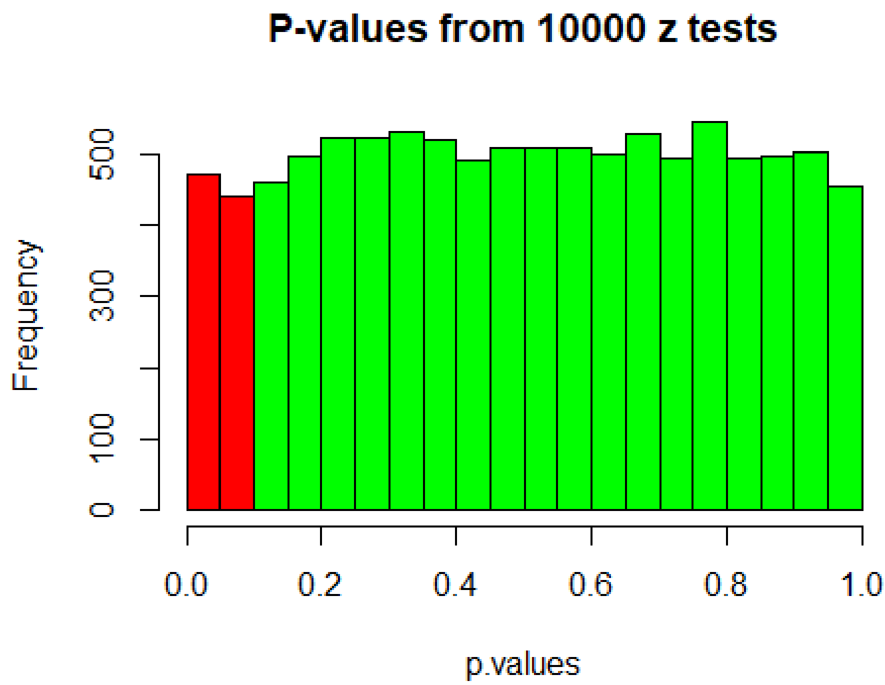
The shape of the simulated distribution strongly resembles a normal distribution that is centered at 399.87 with a standard error of 13.19. The mean and the standard error are very close to the theoretical value with a difference of 0.0009 and -0.2 respectively. Thus, Central Limit Theorem holds.

c.



The z-statistics is normally distributed.

d.




```r
# Proportion of P-values below significance level
Rate = 1-sum(p.values>0.10)/nsim
Rate

## [1] 0.1006
```

The p-values seem to be uniformly distributed. We reject the null hypothesis about 10% of the time. This is a Type I error because we know that the null hypothesis is true.