Stat 3202 Homework 6

Nathan Johnson.9254

2023-04-20

Problem 1:

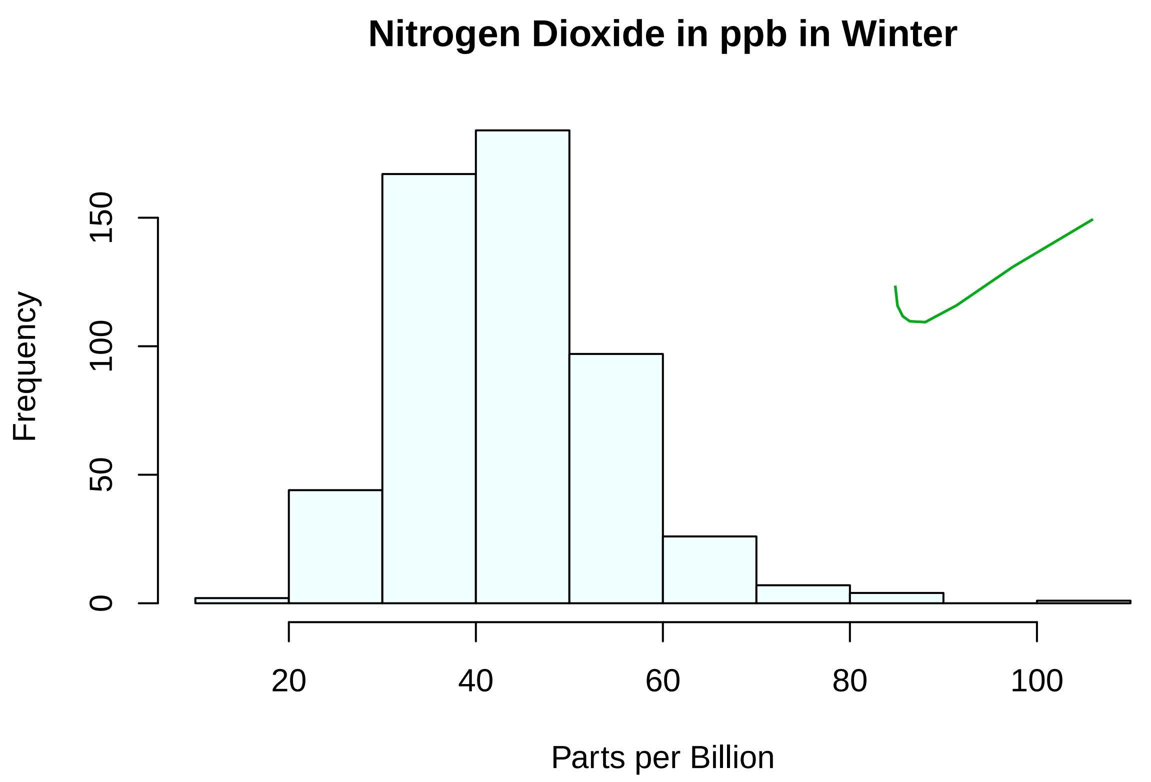
Part a:

```
library(readr)
winter.dataset = read.csv(file = "winter.csv")
```

I was unable to load the data set winter using texmex, but was able to install texmex.

Part b:

```
winter.NO2 = winter.dataset$NO2

hist(winter.NO2, col = "azure", breaks = 10, main = "Nitrogen Dioxide in ppb in Winter", xlab = "Parts p
```



The histogram roughly approximates a normal distribution with a right skewed tail.

Part c:

$H_0 : \mu_{NO_2} = 44$ vs $H_A : \mu_{NO_2} \neq 44$.

Since variance is unknown, we will be using:

$$t = \frac{\bar{x} - 44}{s/\sqrt{n}} \sim t_{n-1}$$

.

To construct the confidence interval we will need to follow:

$$\bar{x} \pm t_{1-\alpha/2,n-1)}\frac{s}{\sqrt{n}}$$

```
alpha = 0.01

n = length(winter.NO2)
s2 = var(winter.NO2)
xbar = mean(winter.NO2)

t = qt(1-alpha/2,n-1)

CIlower = xbar - t*(sqrt(s2)/sqrt(n))
CIupper = xbar + t*(sqrt(s2)/sqrt(n))

c(CIlower,CIupper)
```

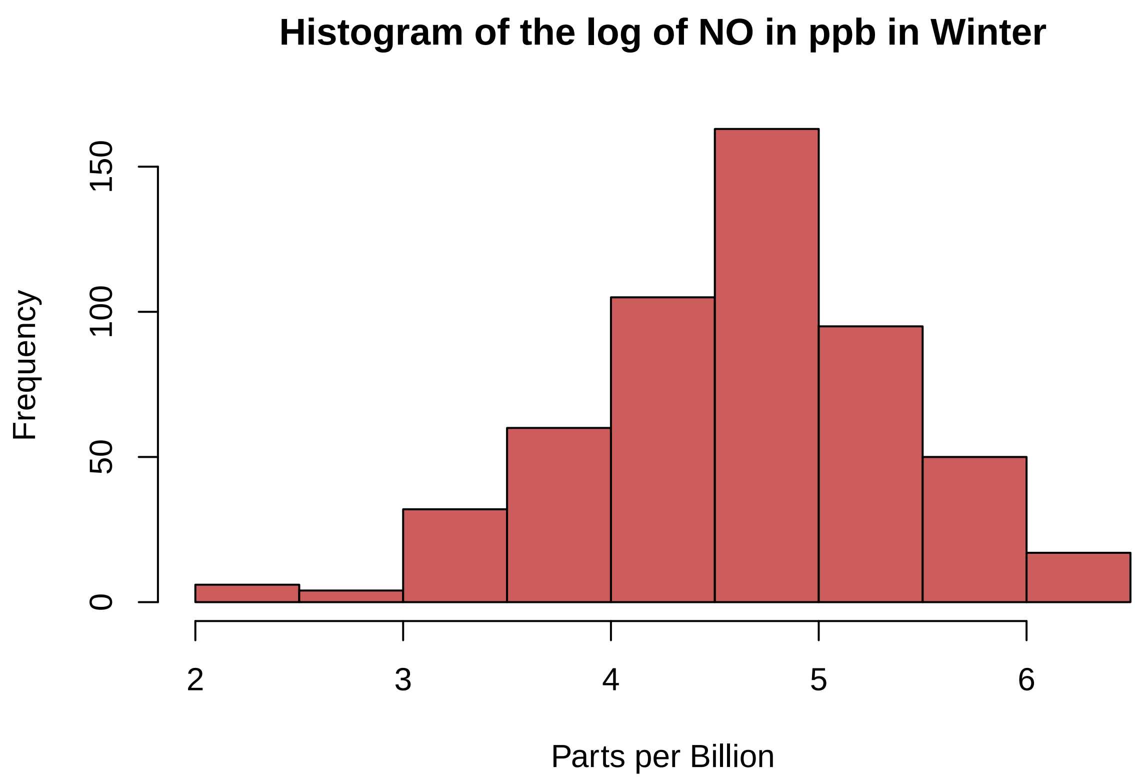
```
## [1] 42.90416 45.43419
```

We are 99% confident the true mean of NO2 in Parts per Billion is between 42.90 and 45.43, thus we fail to reject the null hypothesis.

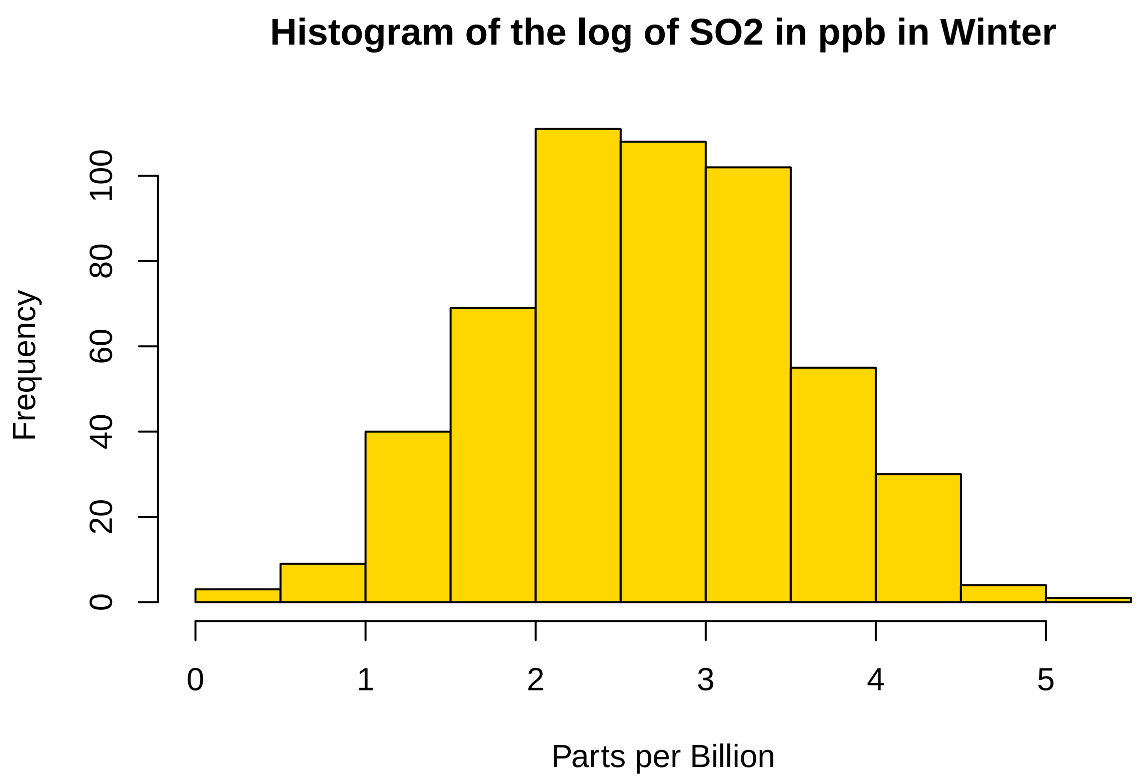
Context ? -1

Part d:

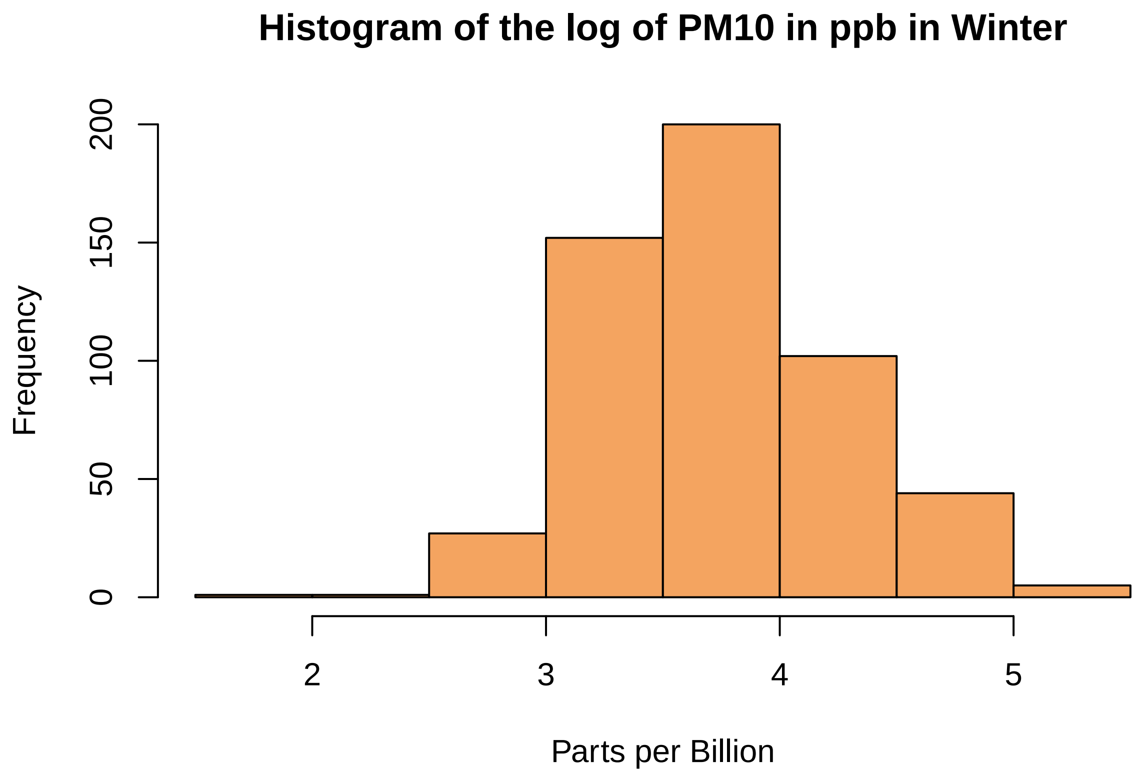
```
winter.NO = winter.dataset$NO
winter.SO2 = winter.dataset$SO2
winter.PM10 = winter.dataset$PM10

hist(log(winter.NO), main = "Histogram of the log of NO in ppb in Winter", xlab = "Parts per Billion", 
```




```
hist(log(winter.SO2), main = "Histogram of the log of SO2 in ppb in Winter", xlab = "Parts per Billion"
```




```
hist(log(winter.PM10), main = "Histogram of the log of PM10 in ppb in Winter", xlab = "Parts per Billio
```

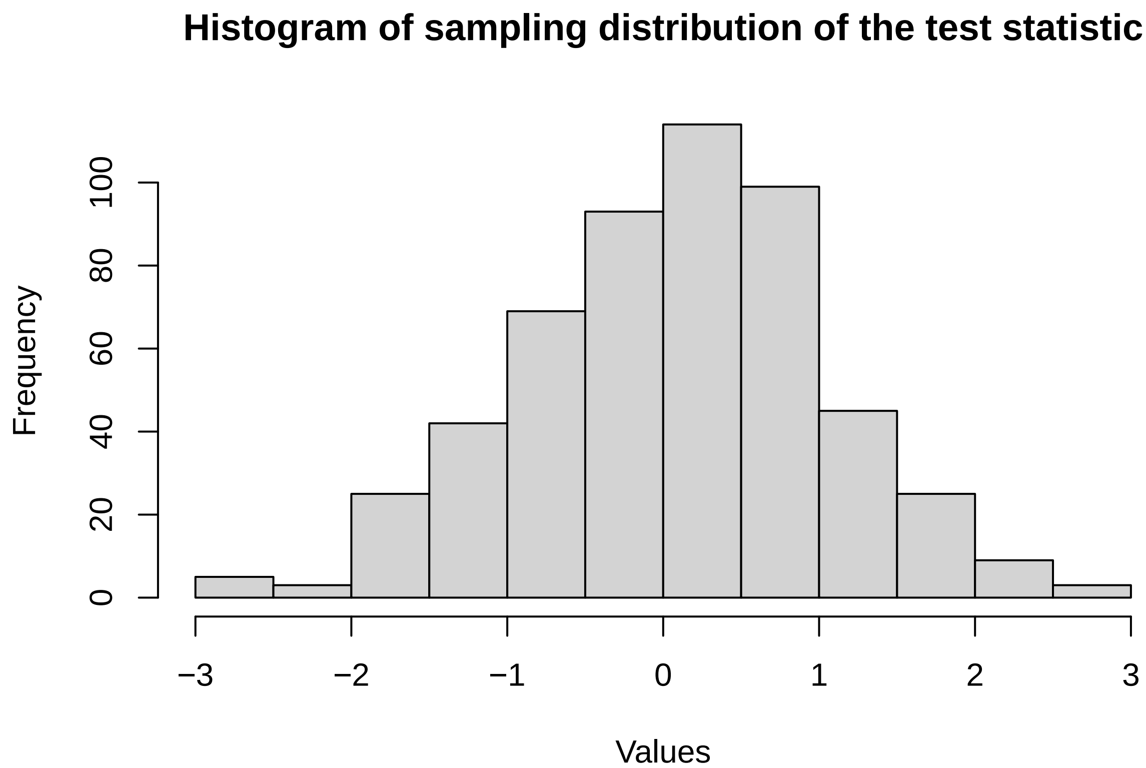


The log of NO histogram and the log of PM10 histogram have slight left and right skews, but not enough to justify using something other than the normal distribution to describe them.

Part e: $H_0 : \mu_{log(SO_2)} = 2.8$ vs $H_A : \mu_{log(SO_2)} < 2.8$.

As in part c, we will also be using the t test statistic.

```
alpha = 0.01

n = length(winter.SO2)
mean.log = mean(log(winter.SO2))
s2.log = var(log(winter.SO2))
mu = 2.8

samp.dist = rt(n, n-1)

hist(samp.dist, main = "Histogram of sampling distribution of the test statistic", xlab = "Values")
```




```
t.stat = (mean.log - mu) / (sqrt(s2.log)/sqrt(n))
t.stat
```

```
## [1] -3.451765
```

```
rr = qt(alpha, n-1)
rr
```

```
## [1] -2.333391
```

```
p = pt(t.stat, n-1, lower.tail = TRUE)
p
```

```
## [1] 0.0003006567
```

```
t = qt(1-alpha, n-1)
CIupper = mean.log + t * (sqrt(s2.log)/sqrt(n))
CIupper
```

```
## [1] 2.757114
```

```
t.test(log(winter.SO2), alternative = "less", mu = mu, conf.level = 1-alpha)
```

```
##
##  One Sample t-test
##
## data:  log(winter.SO2)
## t = -3.4518, df = 531, p-value = 0.0003007
## alternative hypothesis: true mean is less than 2.8
## 99 percent confidence interval:
##      -Inf 2.757114
## sample estimates:
## mean of x
##  2.667636
```

The t test statistic is equal to -3.45 and the rejection region is $(-\infty, -2.33]$. Already, we see that since our test statistic is in the rejection region, we reject the null hypothesis. We also have out p-value is equal to 0.0003007 which is less than the alpha level at 0.01 so we reject the null hypothesis. And we also notice that the confidence interval of $(-\infty, 2.76]$ does not contain the null hypothesis of 2.8 and thus, we reject the null hypothesis and can say that the air quality is acceptable for the concentration of SO2.

Part f:

Since we are using a single variance parameter, we will be using the chi square test statistic:

$$\chi^2 = (n-1)s^2/\sigma^2 \sim^{H_0} \chi^2_{n-1}$$

```
set.seed(1)
alpha = 0.01

n = length(winter.NO2)
mean = mean(winter.NO2)
s2 = var(winter.NO2)

sigma2 = 110

samp.dist = rchisq(n, n-1)

hist(samp.dist, main = "Histogram of the sampling distribution of the test statistic", xlab = "Values")

chi.test = (n - 1)*s2/sigma2
chi.test
```

```
## [1] 614.9525
```

```
chisq = qchisq(1-alpha, n-1)
chisq
```

```
## [1] 609.7395
```

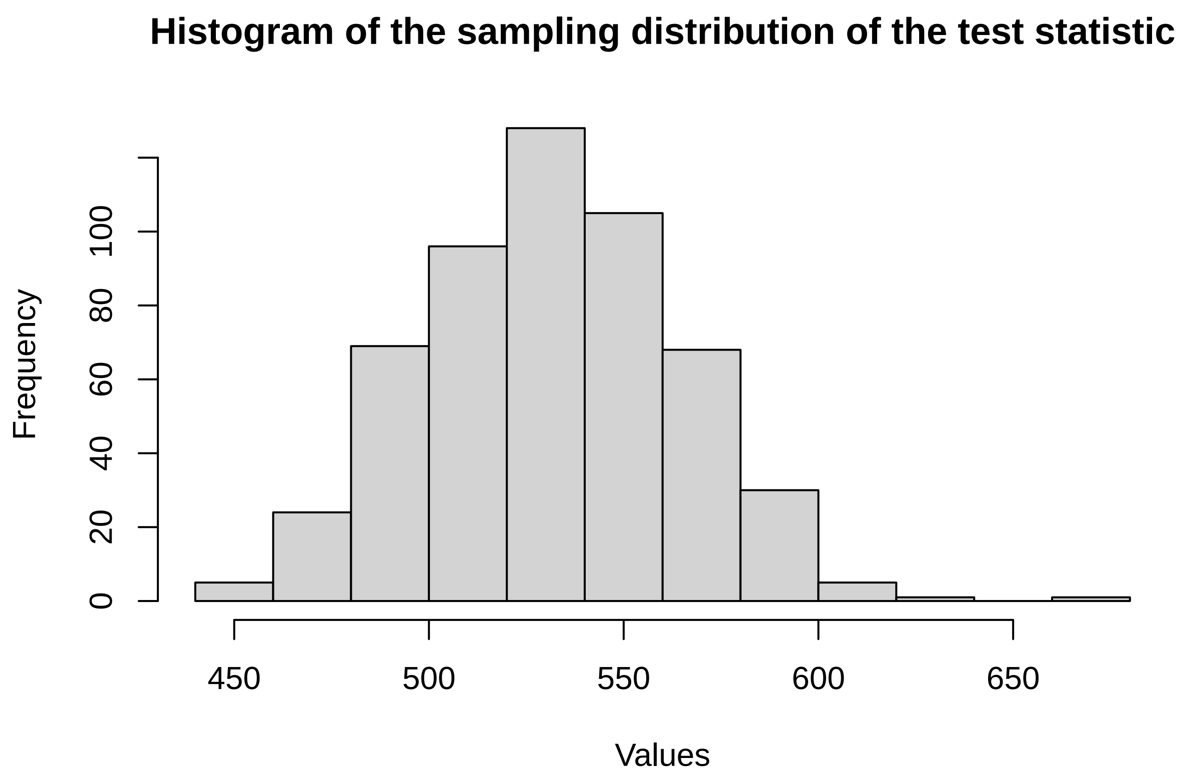
```
chisq.lower = qchisq(1-alpha, n-1)
CIlower = (n-1)*s2/chisq.lower
CIlower
```

```
## [1] 110.9405
```

```r
s2
```

```
## [1] 127.3913
```

```r
library(DescTools)
```




```r
VarTest(winter.NO2, alternative = "greater", sigma.squared = 110, conf.level = 1-alpha)
```

```
##
##  One Sample Chi-Square test on variance
##
## data:  winter.NO2
## X-squared = 614.95, df = 531, p-value = 0.00672
## alternative hypothesis: true variance is greater than 110
## 99 percent confidence interval:
##  110.9405      Inf
## sample estimates:
## variance of x
##      127.3913
```

The test statistic is 614.95 and the rejection region is $[609.74, \infty)$. Since the test statistic is within the rejection region, we reject the null hypothesis in favor of the alternative hypothesis. We can also look at the confidence interval. We are 99% confident the true variance lies within $[110.94, \infty)$. The ~~sample variance~~ lies within the confidence interval which means we fail to reject the null hypothesis according to the confidence interval. Since the p-value is less than the alpha level, according to VarTest, we reject the null hypothesis. I am unsure what this means since I don't believe we were ever taught how to handle contradictory statements, but I am going to assume that since 2 out of the 3 tests reject the null hypothesis, we will reject the null hypothesis in favor of the alternative hypothesis.

-1

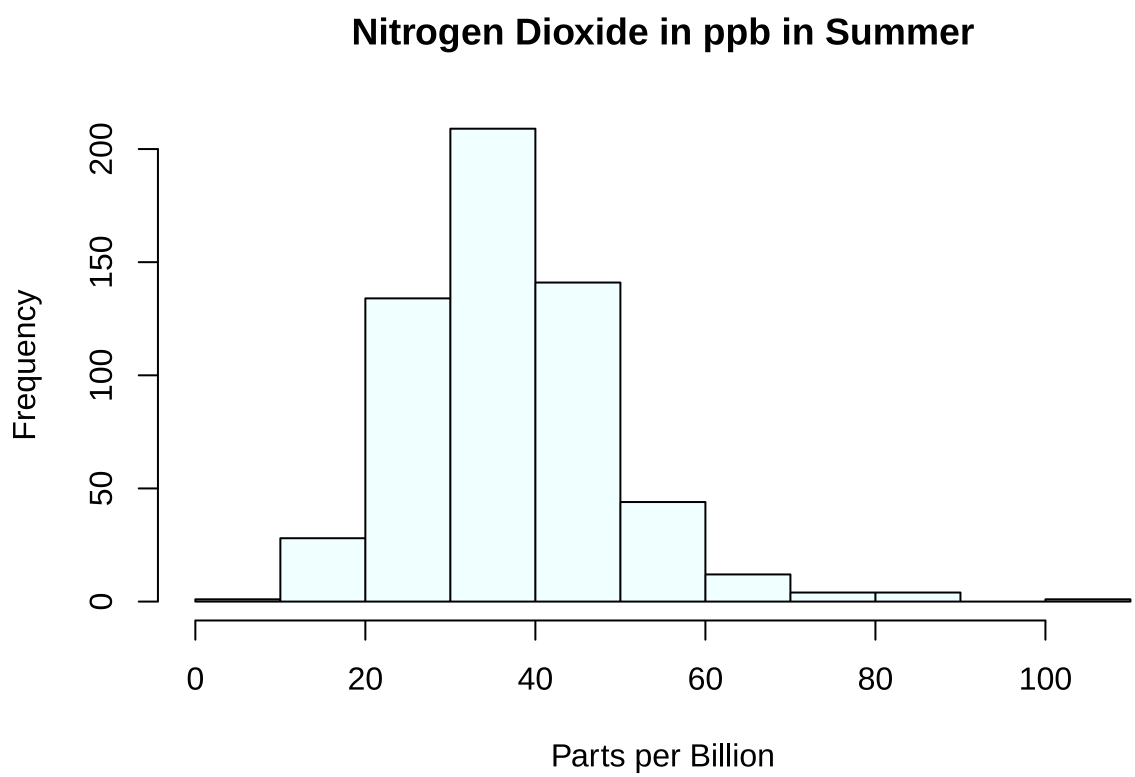
Context?

-1

Problem 2:

Part a:

```
library(readr)
summer.dataset = read.csv(file = "summer.csv")
summer.NO2 = summer.dataset$NO2
hist(summer.NO2, col = "azure", breaks = 10, main = "Nitrogen Dioxide in ppb in Summer", xlab = "Parts 
```

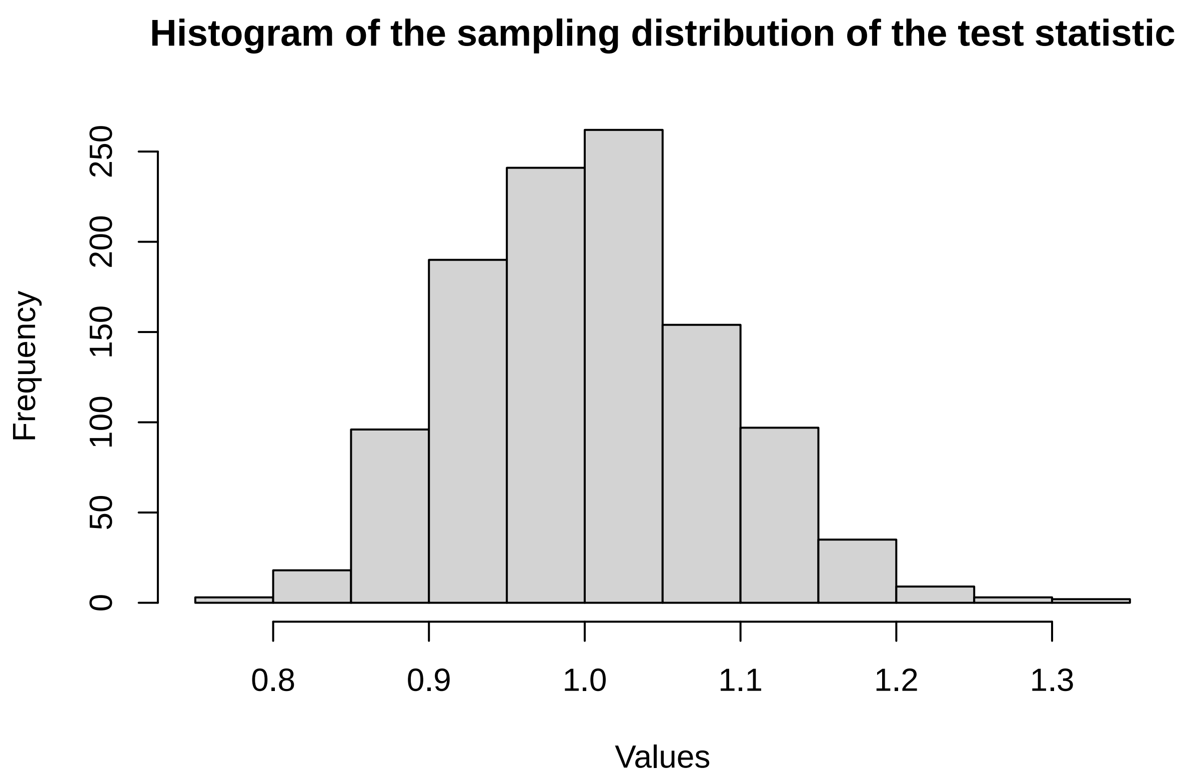


As can be seen, the distribution of NO2 in ppb is roughly normal.

Part b: We need to check: $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ or $H_A : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$.

```
alpha = 0.01

nw = length(winter.NO2)
ns = length(summer.NO2)

s2w = var(winter.NO2)
s2s = var(summer.NO2)
```

```r
samp.dist = rf(nw + ns, nw - 1, ns - 1)

hist(samp.dist, main = "Histogram of the sampling distribution of the test statistic", xlab = "Values")
```




```r
test.stat = s2w / s2s
test.stat
```

```
## [1] 0.9053759
```

```r
F.stat.lower = qf(alpha/2, nw - 1, ns - 1)
F.stat.upper = qf(1-alpha/2, nw - 1, ns - 1)
F.stat.lower
```

```
## [1] 0.8026683
```

```r
F.stat.upper
```

```
## [1] 1.244766
```

```r
CIlower = s2w / (s2s * F.stat.upper)
CIupper = s2w / (s2s * F.stat.lower)
CIlower
```

```
## [1] 0.7273464
```

```
CIupper
```

```
## [1] 1.127958
```

```
VarTest(winter.NO2, summer.NO2, alternative = "two.sided", conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 0.90538, num df = 531, denom df = 577, p-value = 0.2436
## alternative hypothesis: true ratio of variances is not equal to 1
## 99 percent confidence interval:
##  0.7273464 1.1279576
## sample estimates:
## ratio of variances
##          0.9053759
```

The rejection region is $(-\infty, 0.80]$ and $[1.24, \infty)$ and the test statistic is 0.91 which is not contained in the rejection region so we fail to reject the null hypothesis. The confidence interval is $[0.73, 1.13]$ which contains 1 so we fail to reject the null hypothesis once again.

Part c:

Based on part b, the variances of the summer NO2 and winter NO2 in ppb are equal.

```
t.test(winter.NO2, summer.NO2, var.equal = TRUE, conf.level = 1 - alpha)
```

```
##
##  Two Sample t-test
##
## data:  winter.NO2 and summer.NO2
## t = 9.3936, df = 1108, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  4.744404 8.337887
## sample estimates:
## mean of x mean of y
##  44.16917  37.62803
```

We are 99% confident the means of the sample are not equal. Since the summer is lower in NO2 in ppb, it may be less hazardous to individuals with asthma. That said, summer and winter are the same variance so summer still may be hazardous.

Problem 3: Part a: We must do a single population proportion parameter. $np = 75 * (12/75) = 12 > 5$. $n(1-p) = 75 * (63/75) = 63 > 5$. We must test for: $H_0 : p = 0.2$ vs. $H_A : p < 0.2$.

The drug company will need to disclose this side effect if the null hypothesis is not contained in the confidence interval of the population proportion.

Part b:

```r
alpha = 0.05
prop.test(12,75,alternative = "less", p = 0.2,correct = FALSE,conf.level = 1 - alpha)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  12 out of 75, null probability 0.2
## X-squared = 0.75, df = 1, p-value = 0.1932
## alternative hypothesis: true p is less than 0.2
## 95 percent confidence interval:
##  0.0000000 0.2412619
## sample estimates:
##    p
## 0.16
```

Since 0.2 is contained within the confidence interval, the drug company needs to report the side effect in their drug label.