Q #1)

a) numbers = sample(1: nrow(NZBirds.df), size = 35, replace = FALSE)
sample = NZBirds.df[numbers,]

b) mean = 7270.086, and standard deviation = 33754.88.
95 percent confidence interval:
-4325.122 18865.293

c) The possible mean of the body mass for population of these birds lays between 7270 and 33755 grams. We can say this statement with 95% of confidence.

d) Yes. The real mean is 8302.826 which is in the interval. If we take enough samples from the population, the sample behave more like the population. It means that statistics (i.e. sample mean) would have a value closer to the population parameter (i.e. population mean). The interval is distance from point estimate that try to embrace the population parameter.

e)
```
> MuEgg = mean(sampleEgg$Egg)
> MuEgg
[1] 59.51429
> SDEgg = sd(sampleEgg$Egg)
> SDEgg
[1] 44.72423
>
> ciEgg = t.test(sampleEgg$Egg, mu = MuEgg, conf.level = 0.95)$conf.int
> ciEgg
[1] 44.15097 74.87760
attr(,"conf.level")
[1] 0.95
>
mean(c(NZBirds.df$Egg.Length[1:45],NZBirds.df$Egg.Length[48:132]))
[1] 60.65385
```

f) The point estimates of proportion of flightless birds for extinct species is
$P_1 = \frac{y1}{n1} = \frac{21}{38} = 0.5526$
The point estimates of proportion of flightless birds for Non extinct species is
$P_2 = \frac{y2}{n2} = \frac{7}{78} = 0.0897$
For 95% CI , alpha = 1-0.95 = 0.05, alpha/2 = 0.025. So, $Z_{alpa/2} = 1.96$
So, the 95% Ci of difference between the proportions is
$P_1 – P_2 \pm \sqrt{\frac{P1Q1}{n1} - \frac{P2Q2}{n2}} = (0.5526 – 0.0897) \pm \sqrt{\frac{0.5526*0.4474}{38} - \frac{0.0897*0.9103}{78}} = 0.462 \pm 0.1703$
= (0.2926, 0.6332)

g) Since the entire CI is above 0, we can say that we are 95% confident that the proportion of flightless birds is greater for extinct species than for non-extinct species. The CI supports the ecologists’ theory.

Q #2)

```
set.seed(120);x=rnorm(100,mean=1312,sd=422)   # fake data to understand t-tes
ts
set.seed(100); y=rnorm(47,mean=1352,sd=271)

> t.test(x,y,mu =1352 - 1312,var.equal=FALSE,conf.level = 0.90)

        Welch Two Sample t-test

data:  x and y
t = -1.7409, df = 144.995, p-value = 0.08381
alternative hypothesis: true difference in means is not equal to 40
90 percent confidence interval:
 -143.42290   35.38182
sample estimates:
mean of x mean of y 
 1324.452  1378.473 

> t.test(x,y,mu =1352 - 1312,var.equal=TRUE,conf.level = 0.90)

        Two Sample t-test

data:  x and y
t = -1.3758, df = 145, p-value = 0.171
alternative hypothesis: true difference in means is not equal to 40
90 percent confidence interval:
 -167.1498   59.1087
sample estimates:
mean of x mean of y 
 1324.452  1378.473 


> var.test(x, y)

        F test to compare two variances

data:  x and y
F = 4.52, num df = 99, denom df = 46, p-value = 1.396e-07
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.680330 7.279406
sample estimates:
ratio of variances 
          4.519985 
```

Q #3)

$$y \sim N(0, \sigma^2)$$

$$\rightarrow Z \sim N(0,1)$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2 z^2}$$

$$F_X(x) = P(X \leq x)$$

$$= P(Z^2 \leq x)$$

$$= P(-\sqrt{x} \leq z \leq \sqrt{x})$$

$$= F_Z(\sqrt{x}) - F_Z(-\sqrt{x})$$

$$f_X(x) = \frac{1}{2} x^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-1/2 x} +$$

$$\frac{1}{2} x^{-1/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-1/2 x}$$

$$= \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-1/2 x}$$

$$= \Gamma(1/2, 2)$$

$$\therefore \quad x \sim \chi^2$$

Q #4)

a) H0 : μ = 2
   H1 : μ ≠ 2

b) t = -1.02 and p = 0.322

c) $p < 0.05$ is rejection region which is $|t| > t_{0.025}$.

d) $p > 0.05$. There is no evidence to reject null hypothesis.

Q #5)

a)

```
> ci=t.test(WL$DOC,mu=15,conf.level = 0.9)$conf.int
> ci
[1] 10.08013 18.95187
attr(,"conf.level")
[1] 0.9
> pvalue=t.test(WL$DOC,mu=15,conf.level = 0.9)$p.value
> pvalue
[1] 0.8534843
As p-value is greater than 0.05 we accepted the null hypothesis.
```

b) From seeing the 90% CI we can see that a mean of 14grams/m^3 ia within the intercal. So, it is likely to detect a mean of 14grams/m^3.

Q #6)

```
> t.test(fog$RATIO,cl$RATIO,var.equal=FALSE,conf.level = 0.95)

        Welch Two Sample t-test

data:  fog$RATIO and cl$RATIO
t = -1.7459, df = 4.2617, p-value = 0.1514
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.45562346  0.09862346
sample estimates:
mean of x mean of y
  0.27375   0.45225

> pvalue=t.test(fog$RATIO,cl$RATIO,var.equal=FALSE,conf.level = 0.95)$p.value
> pvalue
[1] 0.1513503
```

We can see that p-value is greater than 0.05 so, Null hypothesis is accepted.

That is there is no sufficient evidence that mean of foggy and clear or cloudy ratios are differ at 5%level of significance.

Q #7)

```
> traditional = subset(GT, ENGINE == "Traditional")
> aeroderivative = subset(GT, ENGINE == "Aeroderiv")
```

```
> t.test(traditional$HEATRATE,aeroderivative$HEATRATE,mu = 0,var.equal = TRUE
,conf.level = 0.95)

        Two Sample t-test

data:  traditional$HEATRATE and aeroderivative$HEATRATE
t = -1.2141, df = 44, p-value = 0.2312
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2041.8478   506.5731
sample estimates:
mean of x mean of y
 11544.08  12311.71

> pvalue = t.test(traditional$HEATRATE,aeroderivative$HEATRATE,mu = 0,var.equ
al = TRUE,conf.level = 0.95)$p.value
> pvalue
[1] 0.2311704
```

We can see that p-value is greater than 0.05 for significance level of 0.05, so null hypothesis is accepted. There is no sufficient evidence that variance of these two engine types is different.

```
> advanced = subset(GT, ENGINE == "Advanced")
> t.test(advanced$HEATRATE,aeroderivative$HEATRATE,mu = 0,var.equal = TRUE,co
nf.level = 0.95)

        Two Sample t-test

data:  advanced$HEATRATE and aeroderivative$HEATRATE
t = -4.1945, df = 26, p-value = 0.0002811
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3795.810 -1299.047
sample estimates:
mean of x mean of y
 9764.286 12311.714

> pvalue = t.test(advanced$HEATRATE,aeroderivative$HEATRATE,mu = 0,var.equal
= TRUE,conf.level = 0.95)$p.value
> pvalue
[1] 0.0002810834
```

We can see that null hypothesis is rejected as pvalue is less than 0.05, SO there is a significant evidence that variances of these two engines is different.

Q #8)

```
> DS = subset(GA, Region == "Dry Steppe")
> GD = subset(GA, Region == "Gobi Desert")
> t.test(DS$AntSpecies, GD$AntSpecies,mu=0,var.equal = TRUE, conf.level = 0.9
5)

        Two Sample t-test
```

```
data:  DS$AntSpecies and GD$AntSpecies
t = 0.1821, df = 9, p-value = 0.8595
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -24.74894  29.08228
sample estimates:
mean of x mean of y
 14.00000  11.83333
```

```
> pvalue = t.test(DS$AntSpecies, GD$AntSpecies,mu=0,var.equal = TRUE, conf.le
vel = 0.95)$p.value
> pvalue
[1] 0.8595396
```

a. $H_0 = \frac{variance\ of\ Dry\ Steppe\ region}{variance\ of\ Gobi\ Desert\ region} = 1$. $H_a = \frac{variance\ of\ Dry\ Steppe\ region}{variance\ of\ Gobi\ Desert\ region} \neq 1$.

b. t = 0.1821

c. rejection region = $\propto/2 = 0.05/2 = 0.025.$

d. pvalue = 0.8595

e. We can say that null hypothesis is accepted that is there is no sufficient evidence that variances of above two regions is different.

f. Assumptions:

The data is normal, variances of both regions is equal, samples are independent, standard deviations of both samples are known, the size of both the samples are less than 30.

Q #9)

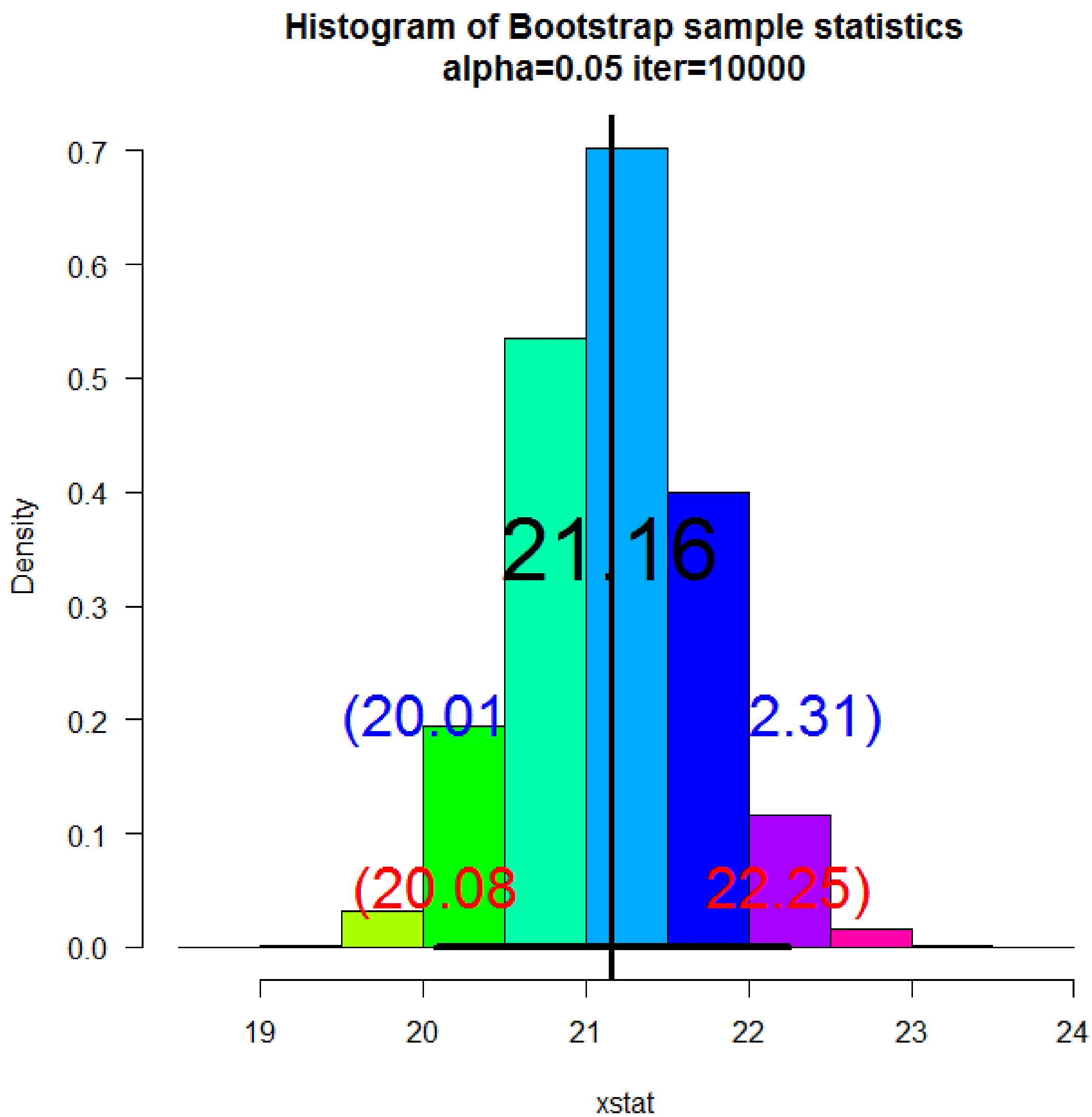
```
> t.test(TP$HUMAN,TP$AUTO,mu=0,var.equal = FALSE)

        Welch Two Sample t-test

data:  TP$HUMAN and TP$AUTO
t = -1.441, df = 13.897, p-value = 0.1717
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -81.06293  15.93793
sample estimates:
mean of x mean of y
 210.8875  243.4500

> pvalue = t.test(TP$HUMAN,TP$AUTO,mu=0,var.equal = FALSE)$p.value
> pvalue
[1] 0.1717404
> pvalue2 = t.test(TP$HUMAN,TP$AUTO,mu=0,var.equal = TRUE)$p.value
> pvalue
[1] 0.1717404
```

From the above, pvalue is greater than 0.05, so null hypothesis is accepted.
There is no significant difference between human and automated thru put.

Q #10)

```
> set.seed(35) # This will give everyone the same sample
> sam=round(rnorm(30,mean=20,sd=3),3)
> sam
```

[1] 23.195 20.399 19.898 19.865 30.014 18.821 21.232 18.313 23.574 21.047 21.535 21.336 17.695 18.497
[15] 14.274 14.664 22.593 18.963 25.515 25.019 22.053 22.871 23.006 23.829 19.038 21.735 21.461 21.659
[29] 21.703 21.049




```
s> obj
$fun
function (x)
mean(x)

$x
 [1] 23.195 20.399 19.898 19.865 30.014 18.821 21.232 18.313 23.574 21.047 21
.535 21.336 17.695 18.497
[15] 14.274 14.664 22.593 18.963 25.515 25.019 22.053 22.871 23.006 23.829 19
.038 21.735 21.461 21.659
```

```
[29] 21.703 21.049

$t
[1] 2.04523

$ci
    2.5%    97.5%
20.08182 22.25104

$cit
[1] 20.01155 22.31198
```