ChickenWeight Assignment-Appendix

Group contribution:

Working alone for pair assignment-1. I have referred few blog and post from stackoverflow and links are mentioned in corresponding sections.For confidence interval calculation "DISCOVERING STATISTICS USING R by Andy Field"(section 2.5.2) book is referred.

Appendix-1

```
library(dplyr)
library(ggplot2)
library(GGally)
library(skimr)
library(gridExtra)
#ChickWeight

df <- as_tibble(ChickWeight)
number_of_observations <- count(df)
number_of_chickens <- length(unique(df$Chick))
number_of_diet<-length(unique(df$Diet))
number_of_days <- max(df$Time)-min(df$Time)
number_of_variables<- length(df)
```

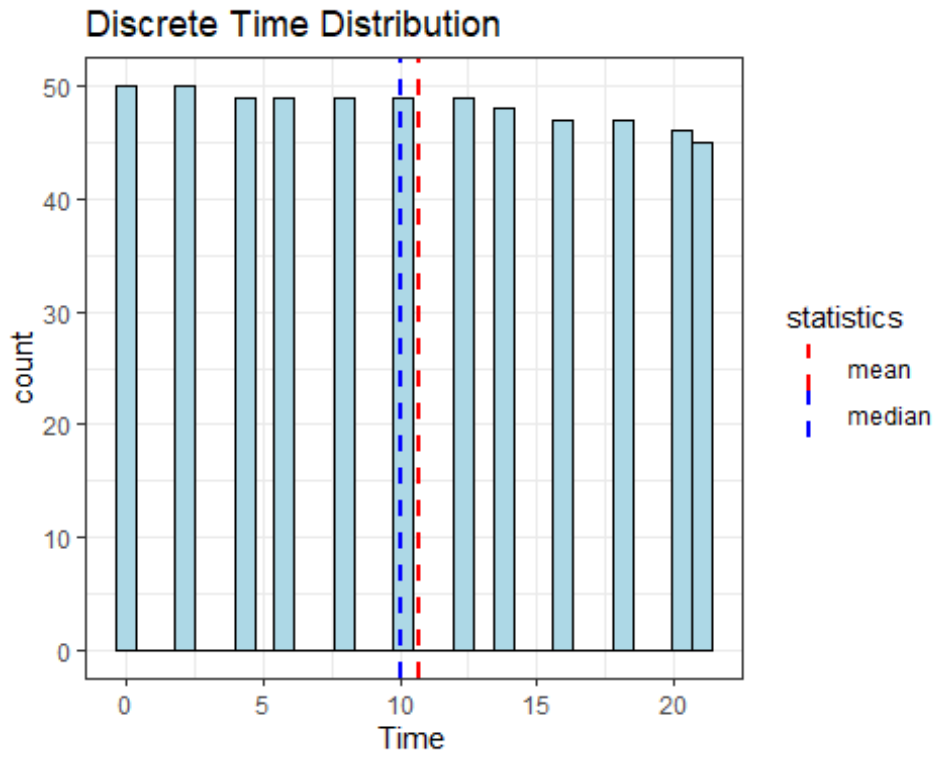
Appendix-2

```
summary(ChickWeight)

##      weight           Time           Chick     Diet
##  Min.   : 35.0   Min.   : 0.00   13     : 12   1:220
##  1st Qu.: 63.0   1st Qu.: 4.00   9      : 12   2:120
##  Median :103.0   Median :10.00   20     : 12   3:120
##  Mean   :121.8   Mean   :10.72   10     : 12   4:118
##  3rd Qu.:163.8   3rd Qu.:16.00   17     : 12
##  Max.   :373.0   Max.   :21.00   19     : 12
##                                  (Other):506

sd(ChickWeight$weight)

## [1] 71.07196
```

Appendix-3

```
ggplot( aes(x = Time), data = df) +
  geom_histogram(color = 'black', fill = 'light blue') +
  geom_vline(aes(xintercept=median(Time),
                 color="median"), linetype="dashed",
             size=1) +
  geom_vline(aes(xintercept=mean(Time),
                 color="mean"), linetype="dashed",
```

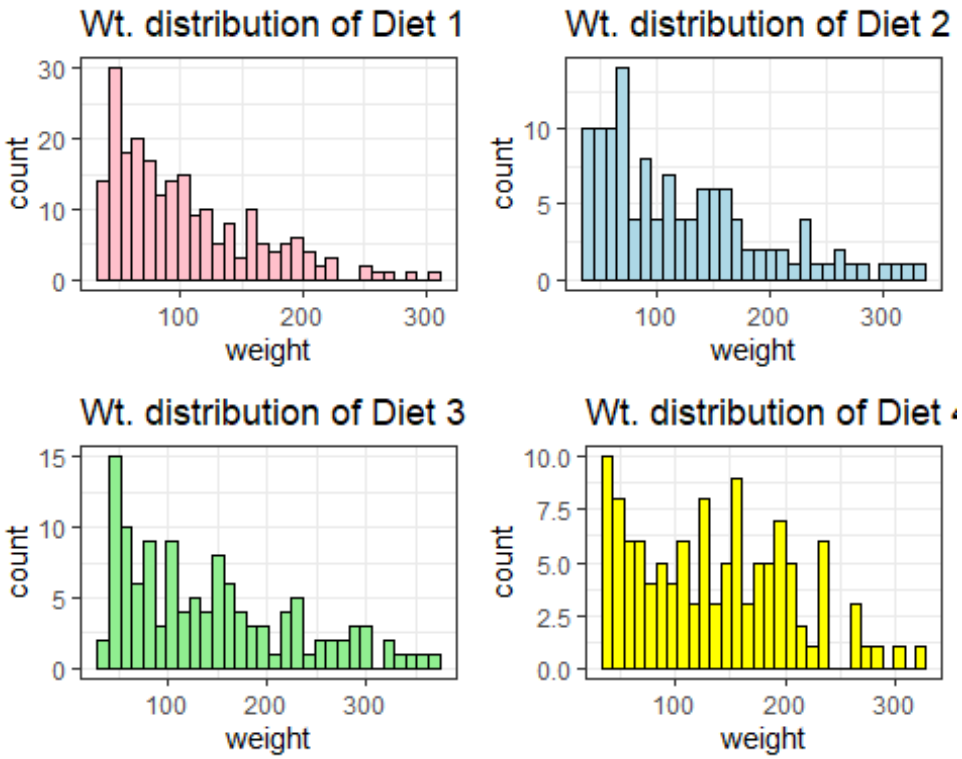
```
            size=1) +
  scale_color_manual(name = "statistics", values = c(median = "blue", mean =
"red"))+ ggtitle("Discrete Time Distribution")+theme_bw()
```



Appendix-4

```
df_diet_1<-df %>% filter(Diet==1)

df_diet_2 <- df %>% filter(Diet==2)
df_diet_3 <- df %>% filter(Diet==3)
df_diet_4 <- df %>% filter(Diet==4)


#par(mfrow=c(2,2))
#plot(  ggplot(data=df_diet_1, aes(x=weight))+ geom_histogram(fill
='pink',color='black'))
#plot(  ggplot(data=df_diet_2, aes(x=weight))+ geom_histogram(fill ='light
blue',color='black'))
#(data=df_diet_3, aes(x=weight))+ geom_histogram(fill ='light
green',color='black')
#ggplot(data=df_diet_4, aes(x=weight))+ geom_histogram(fill
='yellow',color='black')




diet_1 <-ggplot(data=df_diet_1, aes(x=weight))+ geom_histogram(fill
```

```
='pink',color='black')+ggtitle("Wt. distribution of Diet 1")+theme_bw()
diet_2 <-ggplot(data=df_diet_2, aes(x=weight))+ geom_histogram(fill ='light
blue',color='black')+ggtitle("Wt. distribution of Diet 2")+theme_bw()
diet_3 <-ggplot(data=df_diet_3, aes(x=weight))+ geom_histogram(fill ='light
green',color='black')+ggtitle("Wt. distribution of Diet 3")+theme_bw()
diet_4 <-ggplot(data=df_diet_4, aes(x=weight))+ geom_histogram(fill
='yellow',color='black')+ggtitle("Wt. distribution of Diet 4")+theme_bw()

grid.arrange(diet_1,diet_2,diet_3,diet_4,ncol=2,nrow=2)
```



Appendix-5

Note: the degree of freedom for t score = no of samples-1 so we have used qt function for calculation of the same. It accepts the confidence level and degree of freedom as argument and provides corresponding t value. The detailed description for the same is available on: https://stats.stackexchange.com/questions/21596/how-can-i-calculate-a-critical-t-value-using-r

```
day_0 <-  ChickWeight %>% filter(Time==0) %>%group_by(Diet)
day_20 <- ChickWeight %>% filter(Time==20) %>%group_by(Diet)

samples<-function(x) (length(x))


C_I_1 <- function(x) (mean(x)-(1.96*sqrt(var(x)/length(x))))
C_I_2 <- function(x) (mean(x)+(1.96*sqrt(var(x)/length(x))))
```

```
confidence_interval_day_0<- day_0%>%  summarize(mean=mean(weight),
sd=sd(weight),n=samples(weight),C1=C_I_1(weight),C2=C_I_2(weight))

confidence_interval_day_20<- day_20%>%  summarize(mean=mean(weight),
sd=sd(weight),n=samples(weight),C1=C_I_1(weight),C2=C_I_2(weight))
```

Appendix-6

```
confidence_interval_day_0
```

```
## # A tibble: 4 x 6
##   Diet   mean    sd     n    C1    C2
##   <fct> <dbl> <dbl> <int> <dbl> <dbl>
## 1 1      41.4 0.995    20  41.0  41.8
## 2 2      40.7 1.49     10  39.8  41.6
## 3 3      40.8 1.03     10  40.2  41.4
## 4 4      41   1.05     10  40.3  41.7
```

Appendix-7

```
confidence_interval_day_20
```

```
## # A tibble: 4 x 6
##   Diet   mean    sd     n    C1    C2
##   <fct> <dbl> <dbl> <int> <dbl> <dbl>
## 1 1      170.  55.4    17  144.  197.
## 2 2      206.  70.3    10  162.  249.
## 3 3      259.  65.2    10  218.  299.
## 4 4      234.  37.6     9  209.  258.
```