Project-2-cold-storage.R

daoud

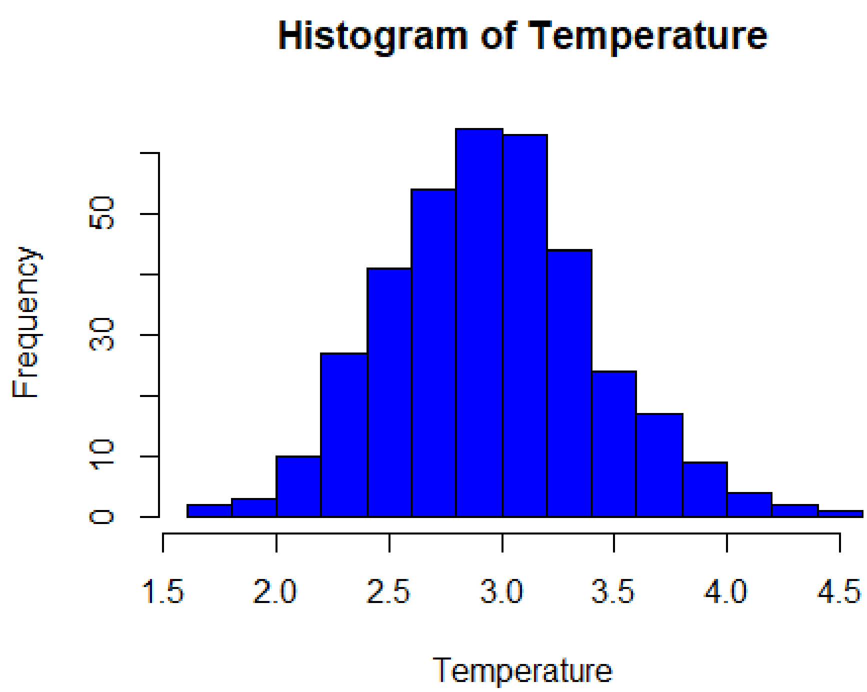
2020-04-10

Problem 1

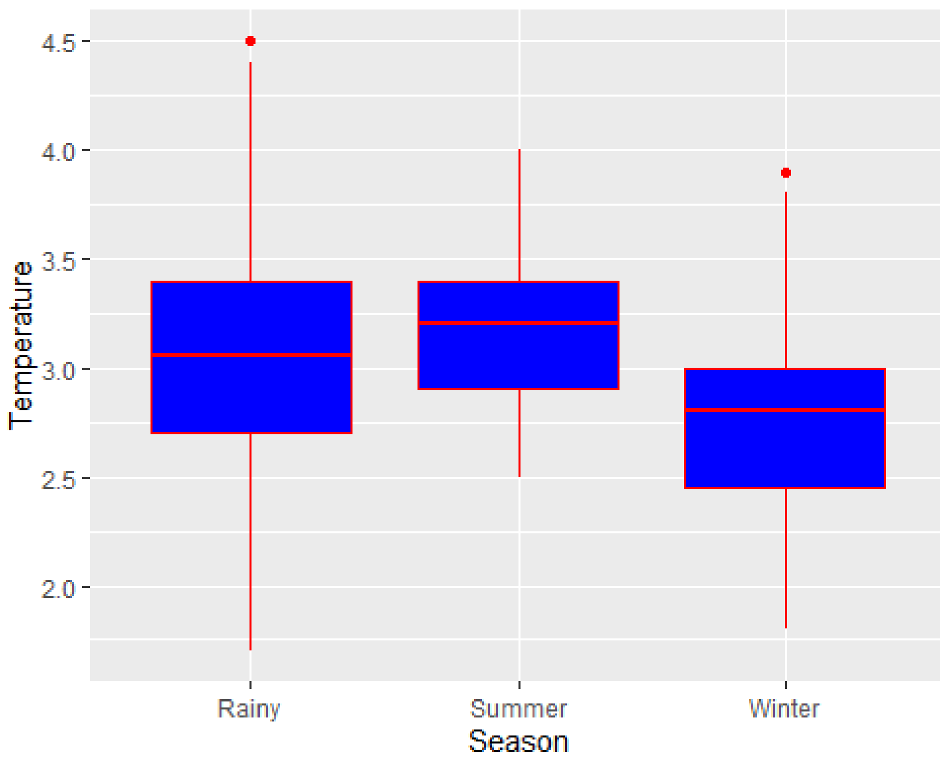
```
install.packages('readr')
install.packages('ggplot2')
install.packages('Rcmdr')
library(readr)

## Warning: package 'readr' was built under R version 3.6.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

library(Rcmdr)

## Warning: package 'Rcmdr' was built under R version 3.6.3

## Loading required package: splines

## Loading required package: RcmdrMisc

## Warning: package 'RcmdrMisc' was built under R version 3.6.3

## Loading required package: car

## Warning: package 'car' was built under R version 3.6.3

## Loading required package: carData

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.6.3

## Loading required package: effects

## Warning: package 'effects' was built under R version 3.6.3

## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

## The Commander GUI is launched only in interactive sessions

##
## Attaching package: 'Rcmdr'

## The following object is masked from 'package:base':
##
##     errorCondition

coldData <- read_csv("C:/Users/daoud/Downloads/PGP DSBA/Fundamental of
Business Statistics/week 4/Cold_Storage_Temp_Data.csv")

## Parsed with column specification:
## cols(
##   Season = col_character(),
##   Month = col_character(),
##   Date = col_double(),
##   Temperature = col_double()
## )

View(coldData)
names(coldData)

## [1] "Season"      "Month"       "Date"        "Temperature"

head(coldData)

## # A tibble: 6 x 4
##   Season Month  Date Temperature
##   <chr>  <chr> <dbl>       <dbl>
## 1 Winter Jan       1         2.3
## 2 Winter Jan       2         2.2
## 3 Winter Jan       3         2.4
## 4 Winter Jan       4         2.8
## 5 Winter Jan       5         2.5
## 6 Winter Jan       6         2.4

dim(coldData)

## [1] 365   4

summary(coldData)

##     Season             Month                Date        Temperature
##  Length:365         Length:365         Min.   : 1.00   Min.   :1.700
##  Class :character   Class :character   1st Qu.: 8.00   1st Qu.:2.700
##  Mode  :character   Mode  :character   Median :16.00   Median :3.000
##                                        Mean   :15.72   Mean   :3.002
##                                        3rd Qu.:23.00   3rd Qu.:3.300
##                                        Max.   :31.00   Max.   :4.500
```

```r
str(coldData)

## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 365 obs. of  4
variables:
##  $ Season     : chr  "Winter" "Winter" "Winter" "Winter" ...
##  $ Month      : chr  "Jan" "Jan" "Jan" "Jan" ...
##  $ Date       : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Temperature: num  2.3 2.2 2.4 2.8 2.5 2.4 2.8 3 2.4 2.9 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Season = col_character(),
##   ..   Month = col_character(),
##   ..   Date = col_double(),
##   ..   Temperature = col_double()
##   .. )

attach(coldData)
hist(Temperature,col = 'blue')
```




```r
ggplot(coldData,aes(y=Temperature,x=Season))+
  geom_boxplot(fill='blue', color="red")
```



observation:

we dont have missing values
our data shape [365 , 4]
we have Season and Month as character , Date and Temperature as numeric
the mean for Temperature is 3.002
possibility of outlier 4.6

Q1 mean cold storage temperature for 3 season:

```
round(tapply(Temperature,Season, mean),2)

##  Rainy Summer Winter
##   3.09   3.15   2.78
```

Q2 overall mean for the full year :

```
total_mean <- mean(Temperature)
total_mean

## [1] 3.002466
```

Q3 Standard Deviation for the full

```
sd<-sd(Temperature)
sd
```

```
## [1] 0.4658319
```

Q4 Assume Normal distribution, probability of temperature having fallen below 2 C :

```
below_2 <- pnorm(2,mean = total_mean,sd = sd,lower.tail = TRUE)
```

Q5 Assume Normal distribution, probability of temperature having going above 4 C :

```
above_4 <-pnorm(4,mean = total_mean,sd = sd,lower.tail =FALSE)
prob2_4 = above_4*100+below_2*100
prob2_4
```

```
## [1] 3.181981
```

Q6 the penalty for the AMC Company:

```
AMC=0.025*100
prob2_4 >= AMC
```

```
## [1] TRUE
```

observation:

as 3.18% > 2.5% ,the probability of temperature going outside the 2 - 4 C during the one-year contract was above 2.5% then the penalty is 10% of AMC.

Q7 one-way ANOVA test to determine if there is a significant difference in Cold Storage:

temperature between rainy, summer and winter seasons and comment on the findings :
H0 : null hypothesis : rainy = summer = winter
Ha : alternative hypothesis : H0 not TRUE

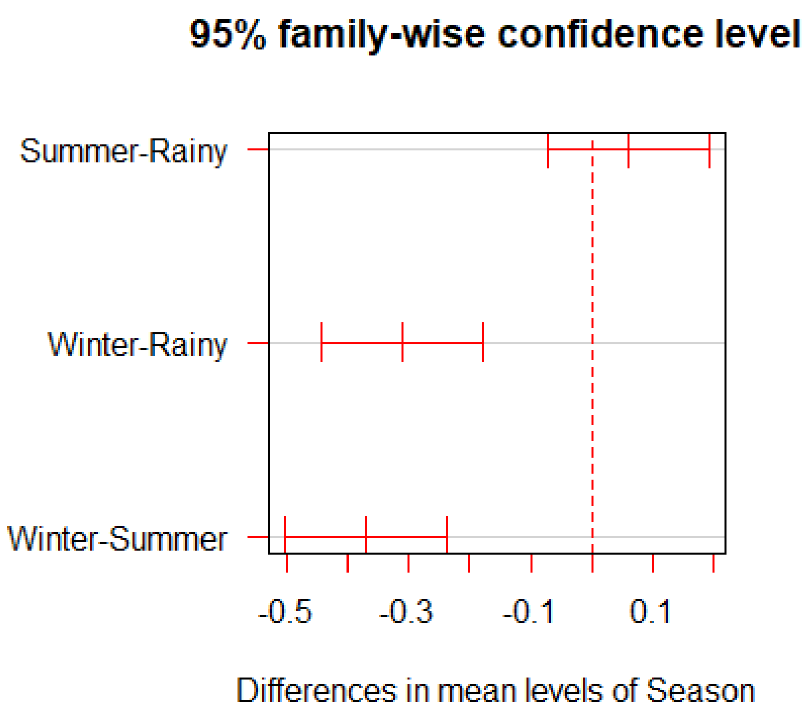
```
cold_anova <- aov(Temperature~Season,data = coldData)
summary(cold_anova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Season        2   9.70   4.848   25.32 5.08e-11 ***
## Residuals   362  69.29   0.191                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

observation:
For the given problem sum of squares due to the factor Season (SSB) is 9.70 and the sum of squares due to error (SSW) is 69.29 .

The total sum of squares (SST) for the data is (9.70+69.29 = 78.99). Since the factor has 3 levels, DF corresponding to Season is 3 - 1 = 2. Total DF is 365 - 1 = 364. Hence DF due to error is 364 - 2 = 362.
Mean sum of squares is obtained by dividing the sums of squares by corresponding DF.
The value of the F-statistic is 25.3 and the p-value is highly significant.
Based on the one way ANOVA we, therefore,
reject the null hypothesis that the three population means are identical.
At least for one Season mean is different from the rest.

```
hypo_ANO <- TukeyHSD(x=cold_anova,conf.level = 0.95)

round(hypo_ANO$Season,2)

##                 diff   lwr   upr p adj
## Summer-Rainy   0.06 -0.07  0.19  0.54
## Winter-Rainy  -0.31 -0.44 -0.18  0.00
## Winter-Summer -0.37 -0.50 -0.24  0.00

par(oma=c(0,7,0,0))
plot(hypo_ANO,las=1, col = "red")
```



observation:

the graphical representation of Temperature comparisons with Season from Tukey's HSD the confident intervals not containing 0 is for the difference between Winter-Rainy and Winter-Summer.
This indicates that population means of these pairs of Season are different.
from the values of the pairwise Summer-Rainy.
it may also be concluded that Temperature from Winter Season is significantly less than the other two.

Problem 2

```
# import data
Cold_Mar2018 <- read_csv("C:/Users/daoud/Downloads/PGP DSBA/Fundamental of
Business Statistics/week 4/Cold_Storage_Mar2018.csv")

## Parsed with column specification:
## cols(
##   Season = col_character(),
##   Month = col_character(),
##   Date = col_double(),
##   Temperature = col_double()
## )

View(Cold_Mar2018)
head(Cold_Mar2018)

## # A tibble: 6 x 4
##   Season Month  Date Temperature
##   <chr>  <chr> <dbl>       <dbl>
## 1 Summer Feb      11         4
## 2 Summer Feb      12         3.9
## 3 Summer Feb      13         3.9
## 4 Summer Feb      14         4
## 5 Summer Feb      15         3.8
## 6 Summer Feb      16         4

summary(Cold_Mar2018)

##     Season             Month                Date       Temperature
##  Length:35          Length:35          Min.   : 1.0   Min.   :3.800
##  Class :character   Class :character   1st Qu.: 9.5   1st Qu.:3.900
##  Mode  :character   Mode  :character   Median :14.0   Median :3.900
##                                        Mean   :14.4   Mean   :3.974
##                                        3rd Qu.:19.5   3rd Qu.:4.100
##                                        Max.   :28.0   Max.   :4.600

dim(Cold_Mar2018)
```

```
## [1] 35  4
```

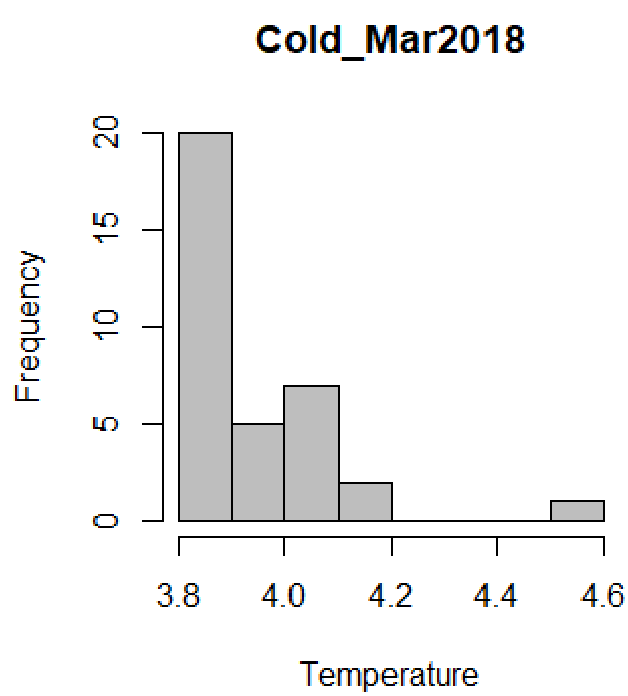
observation:

1. our data is 35 with 4
2.no Missing Data
3.mean of Temperature is 3.974

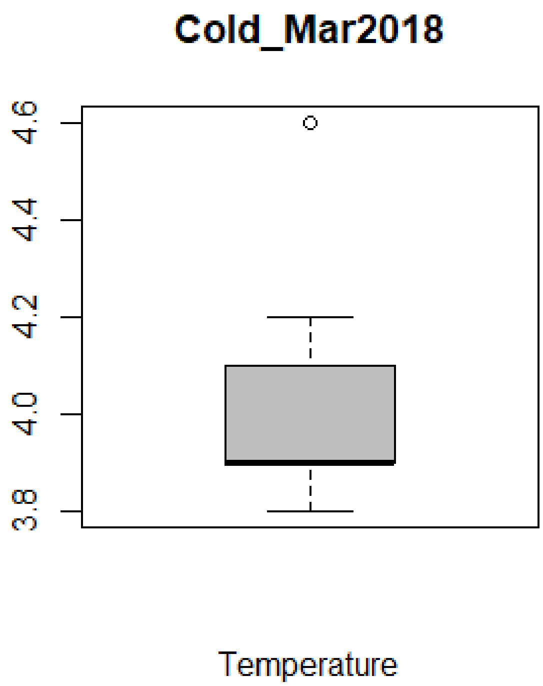
```r
attach(Cold_Mar2018)
```

```
## The following objects are masked from coldData:
##
##     Date, Month, Season, Temperature
```

```r
hist(Temperature,col = 'grey',main = 'Cold_Mar2018',xlab = 'Temperature')
```




```r
boxplot(Temperature,col = 'grey',main = 'Cold_Mar2018',xlab = 'Temperature')
```



Q1 Hypothesis test shall be performed :

since the Supervisor desided temperature as Factor , n > 30 then The size of our sample is 'Large' ,

so we don't need to worry about whether or not the population is normally distributed.

since we don't know the population standard deviation , we concluded that T.test is the need action .

Q2 stat the Hypothesis :

H0: null Hypothesis : ??_Temperature <= 3.9

H1: alternative Hypothesis : H0 not True

lets use T.test with alpha = 0.1 with mean = 3.9 with One tailed test.

```
t.test(Temperature,mu = 3.9,alternative = 'less',conf.level = 0.90)
```

```
##
##  One Sample t-test
##
## data:  Temperature
## t = 2.7524, df = 34, p-value = 0.9953
## alternative hypothesis: true mean is less than 3.9
## 90 percent confidence interval:
##     -Inf 4.00956
```

```
## sample estimates:
## mean of x
##  3.974286
```

observation:

p-value = 0.9953 so p > 0.1 indicates weak evidence against the null hypothesis so we fail to reject the null hypothesis .

Q3 my inference :

since the p-value is large, we fail to reject the null hypothesis
of mean Tempreature less tehn 3.9C and can say that from the last 35 days Tempreature
Cold Storage kept the temperature that agreed to .
about complaints consumers from dairy products mabey due to procurement side from
where Cold Storage is getting the Dairy Products ? or mabey the right Temperature is Not between 2-4C ?