DSBA - Project 2 - Cold Storage Problem

Assessment - Fundamental of Business Statistics

McCombs School of Business - The University of Texas at Austin

Tahmid Bari
PGP - Data Science and Business Analytics

Problem 1

Cold Storage started its operations in Jan 2016. They are in the business of storing Pasteurized Fresh Whole or Skimmed Milk, Sweet Cream, Flavoured Milk Drinks. To ensure that there is no change of texture, body appearance, separation of fats the optimal temperature to be maintained is between 2^0C - 4^0C

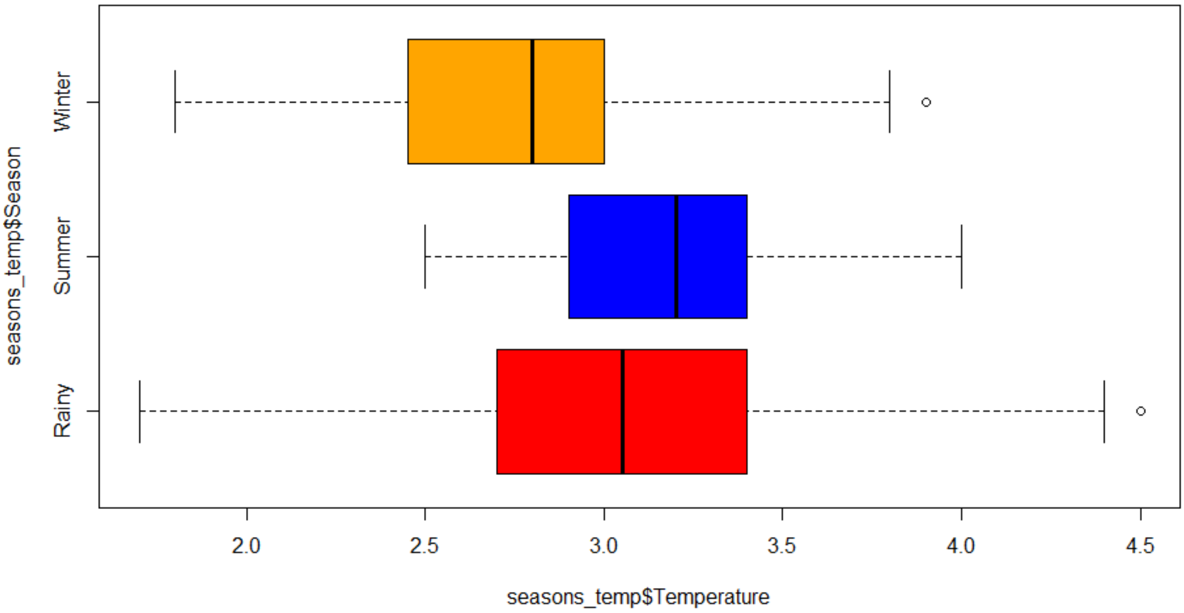
In the first year of business, they outsourced the plant maintenance work to a professional company with stiff penalty clauses. It was agreed that if it was statistically proven that the probability of temperature going outside the 2^0 C - 4^0 C during the one-year contract was above 2.5% and less than 5% then the penalty would be 10% of AMC (annual maintenance contract). In case it exceeded 5% then the penalty would be 25% of the AMC fee. The average temperature data at date level is given in the file "Cold_Storage_Temp_Data.csv"

1. Find mean cold storage temperature for Summer, Winter and Rainy Season

The approach taken in this case is to filter the dataset provided based on seasons i.e. summer, rainy and winter. The package 'dplyr' is being used here for data manipulation and filtering. By using this package, we are able to extract the required columns i.e Seasons and Temperature, filter them by seasons and get a summary of different seasons. The output and the mean cold storage temperature for summer, winter and rainy season are provided below:

```
> summary(winter_temp)
    Season      Temperature
 Rainy  :  0   Min.   :1.800
 Summer:   0   1st Qu.:2.450
 Winter:123    Median :2.800
               Mean   :2.776
               3rd Qu.:3.000
               Max.   :3.900
> summary(summer_temp)
    Season      Temperature
 Rainy  :  0   Min.   :2.500
 Summer:120    1st Qu.:2.900
 Winter:   0   Median :3.200
               Mean   :3.147
               3rd Qu.:3.400
               Max.   :4.000
> summary(rainy_temp)
    Season      Temperature
 Rainy  :122   Min.   :1.700
 Summer:   0   1st Qu.:2.700
 Winter:   0   Median :3.050
               Mean   :3.088
               3rd Qu.:3.400
               Max.   :4.500
```

Therefore, the mean cold storage temperatures for Summer, Winter and Rainy seasons are **3.147^0 C**, **2.776^0 C** and **3.088^0 C** respectively.



Boxplot: from the box plot we see that the temperature between Summer and Rainy season is close

R Code:

```
#===================================================================#
#Data Analysis  -   Cold Storage Problem
#Developer      -   Tahmid Bari
#Date           -   April 11, 2020
#===================================================================#

# Set working directory
setwd("C:/Users/Tahmid Bari/Desktop/Great_Learning/R_Project/Cold_Storage_Problem")
getwd()

# Import the CSV into R
cold_storage_data_temp<-read.csv("Cold_Storage_Temp_Data.csv", header=TRUE)
View(cold_storage_data_temp)

# Install package "dplyr" for data manipulation
install.packages("dplyr")
library("dplyr")

# Create a new data sub-set with the required columns i.e. Season and temperature
seasons_temp<-select(cold_storage_data_temp,Season,Temperature)
View(seasons_temp)

# Filter and view seasons_temp dataset w.r.t winter, summer and rainy
winter_temp<-filter(seasons_temp, Season == "Winter")
summer_temp<-filter(seasons_temp, Season == "Summer")
rainy_temp<-filter(seasons_temp, Season == "Rainy")
View(winter_temp)
View(summer_temp)
```

```
View(rainy_temp)

# Get summary of winter, summary and rainy temperatures
summary(winter_temp)
summary(summer_temp)
summary(rainy_temp)
boxplot(seasons_temp$Temperature~seasons_temp$Season, horizontal = TRUE, col=c("Red",
"Blue", "Orange"))
```

2. Find overall mean for the full year

The overall mean temperature for the full year calculated is **3.002466^{0} C**

R Code:

```
# Overall mean for the full year
mean(cold_storage_data_temp$Temperature)
```

3. Find Standard Deviation for the full year

The standard deviation for the full year calculated is **0. 4658319^{0} C**

R Code:

```
# Standard deviation for the full year
sd(cold_storage_data_temp$Temperature)
```

4. Assume Normal Distribution, what is the probability of temperature having fallen below 2^{0}C?

Since we know (from # 2 and 3) that the mean and standard deviation of temperature throughout the year, we can compute the probability of temperature having fallen below **2^{0} C**. Assuming normal distribution, the probability is **0.01569906** = **1.569%**

R Code:

```
# Probability of temperature having fallen below 2 deg C
mean_temp<-mean(cold_storage_data_temp$Temperature)
sd_temp<-sd(cold_storage_data_temp$Temperature)
X<-2
pnorm(X,mean_temp,sd_temp)
```

5. Assume Normal Distribution, what is the probability of temperature having gone ab over 4^{0} C?

Similar to # 4, since we know the mean and standard deviation of temperature, the probability of temperature having gone above **4^{0} C** is **0.01612075** = **1.612%**

R Code:

```
# Probability of temperature having gone above 4 deg C
Y<-4
prob<-1-pnorm(Y,mean_temp,sd_temp)
prob
```

6. What will be the penalty for the AMC Company?

In this case, we have to compute the probability of temperature going outside 2^0 C - 4^0 C during the one-year contract. Since the temperature record throughout the year follows normal distribution, the probability of temperature going outside 2^0 C - 4^0 C is **0.03181981** = **3.181%**. Since is above 2.5% but below 5%, the **penalty would 10% of the AMC amount**.

R Code:

```
# Penalty for AMC company
Xl<-2
Xu<-4
P_Xl<-pnorm(Xl,mean_temp,sd_temp)
P_Xu<-1-pnorm(Xu,mean_temp,sd_temp)
P_total<-P_Xl+P_Xu
P_total
```

7. Cold Storage temperature between rainy, summer and winter seasons and comment on the findings.

Using the aov() function we can see that the p-value is 5.08e-11 which is really small, which means that we reject the null hypothesis that the 3 means of the temperature for the 3 seasons is equal to each other.

Using the TukeyHSD() function we see that the p-value of the mean between Summer-Rainy season temperature is 0.5376924 ~ 53.77% (confidence level), hence we do not reject the null hypothesis that the means are equal. This implies that there is no significant difference between the temperatures in Summer and Rainy season.

p-value of the means between Rainy-Winter and Winter-Summer is really small, which means there is significant difference in the temperature.

```
> summary(seasons_temp)
    Season     Temperature
 Rainy :122   Min.   :1.700
 Summer:120   1st Qu.:2.700
 Winter:123   Median :3.000
              Mean   :3.002
              3rd Qu.:3.300
              Max.   :4.500
> TukeyHSD(seasons_tempaov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = seasons_temp$Temperature ~ seasons_temp$Season, data = seasons_temp)

$`seasons_temp$Season`
                    diff         lwr        upr     p adj
Summer-Rainy   0.05979508 -0.07258434  0.1921745 0.5376924
Winter-Rainy  -0.31128215 -0.44284519 -0.1797191 0.0000002
Winter-Summer -0.37107724 -0.50318954 -0.2389649 0.0000000
```

R Code:

```
# Perform a 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.
seasons_tempaov = aov(seasons_temp$Temperature~seasons_temp$Season, data = seasons_te
mp)
summary(seasons_temp)
TukeyHSD(seasons_tempaov)
```

Problem 2

In Mar 2018, Cold Storage started getting complaints from their clients that they have been getting complaints from end consumers of the dairy products going sour and often smelling. On getting these complaints, the supervisor pulls out data of the last 35 days' temperatures. As a safety measure, the Supervisor decides to be vigilant to maintain the temperature at 3.9^{0} C or below.

Assume 3.9^{0} C as the upper acceptable value for mean temperature and at alpha = 0.1. Do you feel that there is a need for some corrective action in the Cold Storage Plant or is it that the problem is from the procurement side from where Cold Storage is getting the Dairy Products? The data of the last 35 days is in "Cold_Storage_Mar2018.csv"

1. **Which Hypothesis test shall be performed to check if corrective action is needed at the cold storage plant? Justify your answer.**

H0: $\mu \leq 3.9$ (temperature is 3.9^{0} C or below)
H1: $\mu > 3.9$ (temperature is more than 3.9^{0} C)
Calculate using z-test

Mean (X) = 3.974, Standard deviation (σ) = 0.16
z value = 2.75, for alpha = 0.1 critical z value = 1.28, p-value = 0.002958384 = 0.30%
Since p-value is less than 10% we reject the Null hypothesis (H0: $\mu \leq 3.9$) in favor of the Alternate hypothesis (H1: $\mu > 3.9$). It is likely that corrective action is needed at the cold storage plant.

```
> summary(cold_storage_data_prob2)
    Season     Month         Date         Temperature
 Summer:35   Feb:18    Min.   : 1.0   Min.   :3.800
             Mar:17    1st Qu.: 9.5   1st Qu.:3.900
                       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
> m2 = mean(cold_storage_data_prob2$Temperature)
> m2
[1] 3.974286
> s2 = sd(cold_storage_data_prob2$Temperature)
> s2
[1] 0.159674
> z_cal = (m2 - 3.9)/(s2/sqrt(35))
> z_cal
[1] 2.752359
> pnorm(-abs(z_cal))
[1] 0.002958384
```

R Code:

```
# Which Hypothesis test shall be performed to check if corrective action is needed at the cold stor
age plant
# z-test
cold_storage_data_prob2 = read.csv("Cold_Storage_Mar2018.csv")
summary(cold_storage_data_prob2)

m2 = mean(cold_storage_data_prob2$Temperature)
m2
```

```
s2 = sd(cold_storage_data_prob2$Temperature)
s2

z_cal = (m2 - 3.9)/(s2/sqrt(35))
z_cal
pnorm(-abs(z_cal))
```

2. State the Hypothesis, perform hypothesis test and determine p-value

H0: $\mu \leq 3.9$ (temperature is 3.9^0 C or below)

H1: $\mu > 3.9$ (temperature is more than 3.9^0 C)

Calculate using t-test

t = 2.7524, df = 34, p-value = 0.004711

alternative hypothesis: true mean is greater than 3.9

90 percent confidence interval:

3.939011 Inf

sample estimates:

mean of x 3.974286

Since p-value is less than 10% we reject the Null hypothesis (H0: $\mu \leq 3.9$) in favor of the Alternate hypothesis (H1: $\mu > 3.9$). Hence it proves that corrective action is needed at the cold storage plant.

```
> t.test(cold_storage_data_prob2$Temperature, mu = 3.9, alternative = "greater", conf.level = 0.9)

        One Sample t-test

data:  cold_storage_data_prob2$Temperature
t = 2.7524, df = 34, p-value = 0.004711
alternative hypothesis: true mean is greater than 3.9
90 percent confidence interval:
 3.939011      Inf
sample estimates:
mean of x
 3.974286
```

R Code:

```
# State the Hypothesis, perform hypothesis test and determine p-value
# t-test
t.test(cold_storage_data_prob2$Temperature, mu = 3.9, alternative = "greater", conf.level = 0.9)
pnorm(-abs(z_cal))
```

3. Give your inference

Given the 35 days data and using t-test and z-test it shows that the null hypothesis (temperature is 3.9^0 C or below) is rejected in both cases, as the p-value in both the test has come out to the less than 10%. Hence, it can be believed that corrective actions are required at the cold storage plant.