# *Project - Cold Storage*

Model Report

Table of Contents

# Table of Contents

## 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, Flavored 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 - 4 C.

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 - 4 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

```
> #finding mean cold storage temperature for summer,
winter and Rainy Season
> aggregate(x = Temperature, by = list(Season), FUN = mean)
  Group.1        x
1   Rainy 3.087705
2  Summer 3.147500
3  Winter 2.776423
```

### 2- Find overall mean for the full year

```
> #finding overall mean for the full year
> theMean <- mean(Temperature)
> theMean
[1] 3.002466
```

### 3- Find Standard Deviation for the full year

```
> #finding Standard Deviation for the full year
> theSD <- sd(Temperature)
> theSD
[1] 0.4658319
```

### 4- Assume Normal distribution, what is the probability of temperature having fallen below 2 C?

```
> #the probability of temperature having fallen below 2 C
> pnorm(2, mean=theMean, sd=theSD, lower.tail = T)
[1] 0.01569906
> #calculate percentage
> pnorm(2, mean=theMean, sd=theSD, lower.tail = T) * 100
[1] 1.569906
```

5- Assume Normal distribution, what is the probability of temperature having gone above 4 C?

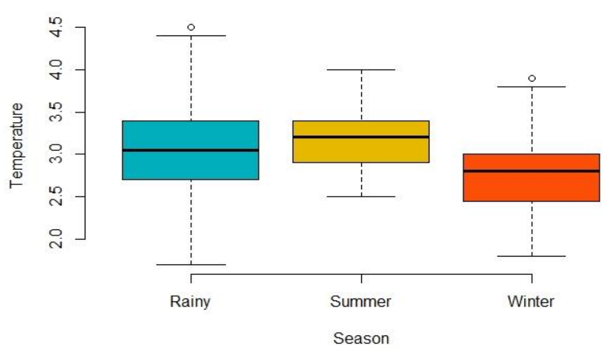
```
> #the probability of temperature having gone above 4 C
> pnorm(4, mean=theMean, sd=theSD, lower.tail = F)
[1] 0.01612075
> #calculate percentage
> pnorm(4, mean=theMean, sd=theSD, lower.tail = F) * 100
[1] 1.612075
```

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

While the probability of temperature going outside the 2 - 4 C during the one-year contract is equal to **1.569906 + 1.612075= 3.181981 %, which is above 2.5% and less than 5%,** so the penalty would be <u>10% of AMC (annual maintenance contract).</u>

7- 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.

```
> group_by(Cold_Storage_Temp_Data, Season) %>%
+     summarise(
+         count = n(),
+         mean = mean(Temperature, na.rm = TRUE),
+         sd = sd(Temperature, na.rm = TRUE)
+     )
# A tibble: 3 x 4
  Season count  mean    sd
  <fct>  <int> <dbl> <dbl>
1 Rainy    122  3.09 0.527
2 Summer   120  3.15 0.352
3 Winter   123  2.78 0.414
```




From pic above, we can see there is a deferent in groups mean, but we are going to test it to identify is it Significant or not.

The null hypothesis (H0) is that there is no difference among group means. The alternate hypothesis (Ha) is that at least one group differs significantly from the overall mean of the dependent variable.

```
> # Compute the analysis of variance
> one.way <- aov(Temperature ~ Season, data = Cold_Storage_Temp_Data)
> # Summary of the analysis
> summary(one.way)
             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
```

- **The F-value** are equal to **4.848**.
- **The Pr(>F)** are equal to **0.0051**.

Because the ***p*-value** of the independent variable season, which is ≃ **0.0051** is significant ($p < 0.05$), it is likely that season does have a significant effect on average temperature.

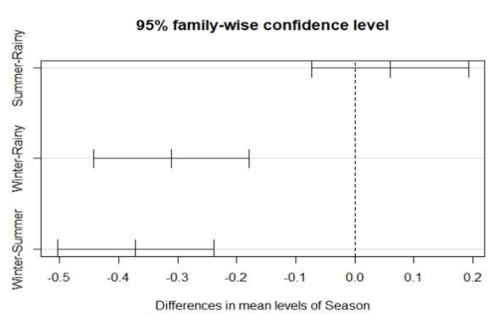
Now we are interested in that the treatment levels differ from one another, hence we are performing a TukeyHSD:

```
> #find how the treatment levels differ from one another
> TukeyHSD(one.way)
  Tukey multiple comparisons of means
    95% family-wise confidence level

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

$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
```

The pairwise comparisons show that Winter has a significantly higher mean yield than both Rainy and Summer, but the difference between the mean yields of Rainy and Summer is not statistically significant.
A Tukey post-hoc test revealed significant pairwise differences between Winter-Summer, with an average difference of -0.4 C in ($p = 0$) and between Winter-Rainy, with an average difference of -0.3 C in ($p \simeq 0$), and between Summer-Rainy, with an average difference of 0.059 C in ($p < 0.54$).




Therefore, we can say that the Winter has statistical significant impact on temperature.

## 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 C or below.

Assume 3.9 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.

While we do not know the population standard deviation, so we have to estimate it with the sample standard deviation.

Therefore, the test we need to use is a one-sample t-test for means.

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

We assume that:

- The dependent variable must be continuous.
- The observations are independent of one another.
- The dependent variable should be approximately normally distributed.
- The dependent variable should not contain any outliers.

The one sample t-test determines whether the sample mean is statistically different from a known or hypothesized population mean.

So that we are going to use one sample right-tiled t-test for mean to check the temperature mean is it exceed the upper acceptable degree, which is 3.9C .

It should be like:

- Null Hypothesis: The sample mean is less than or equal to 3.9C
- Hypothesis: The sample mean is greater than 3.9C

$$H_0: \mu \leq 3.9C \text{ vs. } H_a: \mu > 3.9C$$

To perform this test we are going to use **t.test** function with this parameter:

- X = Cold_Storage_Mar2018$Temperature Perform
- mu = 3.9
- alt = 'greater'

```
> t.test(x = Cold_Storage_Mar2018$Temperature, mu = 3.9, alt = 'greater')

	One Sample t-test

data:  Cold_Storage_Mar2018$Temperature
t = 2.7524, df = 34, p-value = 0.004711
alternative hypothesis: true mean is greater than 3.9
95 percent confidence interval:
 3.928648      Inf
sample estimates:
mean of x
 3.974286
```

From result, we can see that the *p*-value = 0.004711

## 3- Give your inference

Whereas the *p*-value is less than 0.1; therefore, we reject the null hypothesis, and support the alternative hypothesis, so the problem is from Cold Storage Plant.

## Appendix A – Source Code

```
#========================================================================
#
# Project 2 - Cold Storage
#
#========================================================================
#calling all libraries that we are going to use
library(readr)
library(dplyr)
library(ggpubr)
#========================================================================
# problem 1
#========================================================================
#setting up working directory
setwd("………/PGP DSBA/DSBA/Data/Project 2 - Cold Storage")

#reading data from csv file to Cold_Storage_Temp_Data variable and view it
Cold_Storage_Temp_Data <- read.csv("Cold_Storage_Temp_Data.csv")
View(Cold_Storage_Temp_Data)
#check if ther is any NA value in dataset
anyNA(Cold_Storage_Temp_Data)
#Return a summary of the dataset variables.
summary(Cold_Storage_Temp_Data)
#Retrieve the dimension of an object.
dim(Cold_Storage_Temp_Data)
#Get the names of an object.
names(Cold_Storage_Temp_Data)
#Display the internal structure of an dataset.
str(Cold_Storage_Temp_Data)
#Returns the first 10 rows of the dataset.
head(Cold_Storage_Temp_Data, 10)
#Returns the last 10 rows of the dataset.
tail(Cold_Storage_Temp_Data, 10)
#objects in the dataset can be accessed by simply giving their names
attach(Cold_Storage_Temp_Data)
#findoing mean cold storage temperature for summer,winter and Rainy Season
aggregate(x = Temperature,                # Specify data column
          by = list(Season),              # Specify group indicator
          FUN = mean)                     # Specify function (i.e. mean)
#Finding overall mean for the full year
theMean <- mean(Temperature)
#Finding Standard Deviation for the full year
theSD <- sd(Temperature)
#the probability of temperature having fallen below 2 C
pnorm(2, mean=theMean, sd=theSD, lower.tail = T)
#calculate percentage
pnorm(2, mean=theMean, sd=theSD, lower.tail = T) * 100
#the probability of temperature having gone above 4 C
pnorm(4, mean=theMean, sd=theSD, lower.tail = F)
#calculate percentage
pnorm(4, mean=theMean, sd=theSD, lower.tail = F) * 100
# one-way ANOVA
#ompute summary statistics by groups - count, mean, sd:
group_by(Cold_Storage_Temp_Data, Season) %>%
  summarise(
```

```
    count = n(),
    mean = mean(Temperature, na.rm = TRUE),
    sd = sd(Temperature, na.rm = TRUE)
  )
# Compute the analysis of variance
one.way <- aov(Temperature ~ Season, data = Cold_Storage_Temp_Data)
# Summary of the analysis
summary(one.way)
#find how the treatment levels differ from one another
TukeyHSD(one.way)
#ploting TukeHSD
plot(TukeyHSD(one.way))
#========================================================================
# problem 2
#========================================================================
#reading data from csv file to Cold_Storage_Mar2018 variable and view it
Cold_Storage_Mar2018 <- read_csv("Cold_Storage_Mar2018.csv")
View(Cold_Storage_Mar2018)
#check if ther is any NA value in dataset
anyNA(Cold_Storage_Mar2018)
#take look to data throgh boxplot
boxplot(Cold_Storage_Mar2018$Temperature)
#performing t-test
t.test(x=Cold_Storage_Mar2018$Temperature, mu = 3.9, alt = 'greater')
#========================================================================
#
# T H E - E N D
#
#========================================================================
```