Great Lakes Institute of Management

MINI PROJECT – COLD STAORAGE CASE STUDY

Contents

1. Project Objective:

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 deg - 4 deg 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 the it was statistically proven that probability of temperature going outside the 20 - 40 C during the one-year contract was above 2.5% and less than 5% then the penalty would be 10% of AMC (annual maintenance case). 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".
The objectives for this particular part is basically a step wise approach to find the normal distribution and give our conclusion if penalty is to be levied or not.

The next part is based on a 2018 fact where there have been temperature fluctuation issues resulting in harming the products stored in the storage.
The questions include: -

- Find mean cold storage temperature for Summer, Winter and Rainy Season
- Find overall mean for the full year
- Find Standard Deviation for the full year
- Assume Normal distribution, what is the probability of temperature having fallen below 2 deg C
- Assume Normal distribution, what is the probability of temperature having gone above 4 deg C
- What will be the penalty for the AMC Company
- State the Hypothesis, do the calculation using z test
- State the Hypothesis, do the calculation using t test
- Give your inference after doing both the tests

Sample Data Set:

	A	B	C	D
1	Season	Month	Date	Temperature
2	Winter	Jan	1	2.4
3	Winter	Jan	2	2.3
4	Winter	Jan	3	2.4
5	Winter	Jan	4	2.8
6	Winter	Jan	5	2.5
7	Winter	Jan	6	2.4
8	Winter	Jan	7	2.8
9	Winter	Jan	8	2.3
10	Winter	Jan	9	2.4
11	Winter	Jan	10	2.8
12	Winter	Jan	11	2.4
13	Winter	Jan	12	2.5
14	Winter	Jan	13	2.6
15	Winter	Jan	14	2.8
16	Winter	Jan	15	3.4
17	Winter	Jan	16	3.9
18	Winter	Jan	17	3.3
19	Winter	Jan	18	3.3
20	Winter	Jan	19	2.8
21	Winter	Jan	20	2.4
22	Winter	Jan	21	2.5
23	Winter	Jan	22	2.3
24	Winter	Jan	23	2.7
25	Winter	Jan	24	2.4
26	Winter	Jan	25	3.5
27	Winter	Jan	26	2.5
28	Winter	Jan	27	2.6
29	Winter	Jan	28	2.8
30	Winter	Jan	29	3.1
31	Winter	Jan	30	2.5
32	Winter	Jan	31	2.4

Cold_Storage_Temp_Data.

	A	B	C	D
1	Season	Month	Date	Temperature
2	Summer	Feb	11	4
3	Summer	Feb	12	3.9
4	Summer	Feb	13	3.9
5	Summer	Feb	14	4
6	Summer	Feb	15	3.8
7	Summer	Feb	16	4
8	Summer	Feb	17	4.1
9	Summer	Feb	18	4
10	Summer	Feb	19	3.8
11	Summer	Feb	20	3.9
12	Summer	Feb	21	3.9
13	Summer	Feb	22	4.6
14	Summer	Feb	23	4.1
15	Summer	Feb	24	4.1
16	Summer	Feb	25	3.9
17	Summer	Feb	26	3.8
18	Summer	Feb	27	3.8
19	Summer	Feb	28	3.9
20	Summer	Mar	1	3.9
21	Summer	Mar	2	3.9
22	Summer	Mar	3	3.9
23	Summer	Mar	4	4.1
24	Summer	Mar	5	3.9
25	Summer	Mar	6	3.9
26	Summer	Mar	7	4.1
27	Summer	Mar	8	4
28	Summer	Mar	9	4.1
29	Summer	Mar	10	3.9
30	Summer	Mar	11	4.1
31	Summer	Mar	12	3.8
32	Summer	Mar	13	4.2

Cold_Storage_Mar2018

2. Assumptions

The sample size of the data set should be greater than 30. According to Central Limit Theorem, irrespective of the original population distribution, the sampling distribution of the mean will approach to a normal distribution as the size of the sample increases and becomes large (>30).
Alpha if not given will be taken as 0.05 but here alpha is provided as 0.1.
We also assume that temperature remains uniform across the day and also the temperature of the storage items is between 2 to 4 degree when it came to the facility. The exterior conditions will have no bearing on the cold storage facility.

3. Exploratory Data Analysis – Step by step approach

1. Environment Set up and Dataset Import
2. Variable Identification
3. Univariate Descriptive Analysis
4. Hypothesis Testing (z-test & t-test)
5. Conclusions

3.1 Environment Set up and Data Import

3.1.1 Install necessary Packages and Invoke Libraries

1. library(readr)
2. library(ggplot2)
3. library(TeachingDemos)

3.1.2 Set up working Directory

Setting working directory to the directory where I have saved the dataset on my local machine to load the data easily.

```
setwd("~/Desktop/PGP-BABI")
```

3.1.3 Import and Read the Dataset

The data set is imported to studio using read.csv command.

```
cold_storage_data = read.csv("Cold_Storage_Temp_Data.csv")
```

3.2 Variable Identification

1. cold_storage_data – to store the dataset of the cold storage from January 2016.
2. cold_march – to store the dataset for March 2018.
3. histinfo – to store the information of the histogram created with the temperature variations.
4. min_temp – to store the minimum temperature of the January 2016 dataset
5. max_temp – to store the maximum temperature of the January 2016 dataset
6. histinfo_march – to store the information of the histogram created with the temperature variations.
7. sd_yearly_temp – to store the yearly standard deviations of the cold_storage_data for the temperature variable.
8. mean_yearly_temp – to store the mean of the temperature variable of the cold_storage_data.

For the entire code refer to Appendix section at the end of the documentation.

Also some basic statistics information include:-

1. summary(cold_storage_data)

```
    Season        Month           Date          Temperature
 Rainy :122   Aug    : 31   Min.   : 1.00   Min.   :1.700
 Summer:120   Dec    : 31   1st Qu.: 8.00   1st Qu.:2.500
 Winter:123   Jan    : 31   Median :16.00   Median :2.900
              Jul    : 31   Mean   :15.72   Mean   :2.963
              Mar    : 31   3rd Qu.:23.00   3rd Qu.:3.300
              May    : 31   Max.   :31.00   Max.   :5.000
              (Other):179
>
```

2. str(cold_storage_data)

```
'data.frame':   365 obs. of  4 variables:
 $ Season     : Factor w/ 3 levels "Rainy","Summer",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Month      : Factor w/ 12 levels "Apr","Aug","Dec",..: 5 5 5 5 5 5 5 5 5 5 ...
 $ Date       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Temperature: num  2.4 2.3 2.4 2.8 2.5 2.4 2.8 2.3 2.4 2.8 ...
>
```

3. summary(cold_march)

```
     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
>
```

4. str(cold_march)

```
'data.frame':   35 obs. of  4 variables:
 $ Season     : Factor w/ 1 level "Summer": 1 1 1 1 1 1 1 1 1 1 ...
 $ Month      : Factor w/ 2 levels "Feb","Mar": 1 1 1 1 1 1 1 1 1 1 ...
 $ Date       : int  11 12 13 14 15 16 17 18 19 20 ...
 $ Temperature: num  4 3.9 3.9 4 3.8 4 4.1 4 3.8 3.9 ...
>
```

Different R functions used:-

1. read.csv () to load the data set
2. summary () to get summary of dataset
3. mean () to get mean of data set
4. sd () to get Standard Deviation of the data set
5. pnorm () to get the probability of the data set

3.3 Descriptive Analysis (Univariate):

In this step the features of the dataset are explored in details. This step would help us get meaningful insights and summaries about the datasets we will be using. The various methods used to gain the insights have been displayed well in section 3.4 of the documentation. Below is the list of tasks that have been performed over the datasets.

Measures of Central Tendency	Measures of Dispersion	Visualization Method
Mean	Range	Histogram
Median	1st Quartile	Boxplot
Mode	3rd Quartile	
Minimum	Inter Quartile Range (IQR)	
Maximum	Variance	
	Standard Deviation	

Cold_Storage_Temp_Data

	Full year	Rainy	Summer	Winter
Min.	1.7	1.7	2.5	1.9
1st Qu.	2.5	2.5	2.8	2.4
Median	2.9	2.9	3.2	2.6
Mean	2.963	3.039	3.153	2.701
3rd Qu.	3.3	3.3	3.4	2.9
Max.	5	5	4.1	3.9

Standard Deviation= 0.51
Probability below 2-degree temperature = 0. 02989406
Probability above 4-degree temperature = 0. 02071425

Cold_Storage_Mar2018

Min.	3.8
1st Qu.	3.9
Median	3.9
Mean	3.974
3rd Qu.	4.1
Max.	4.6

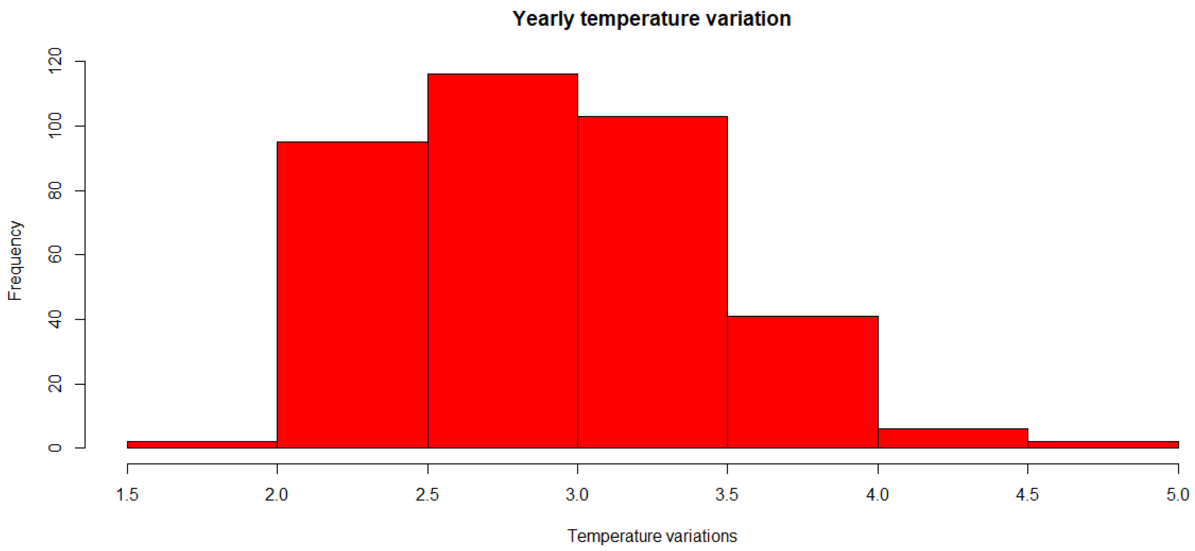
Standard Deviation = 0.159674

3.4 Data Visualisations:

Data Visualisations :-

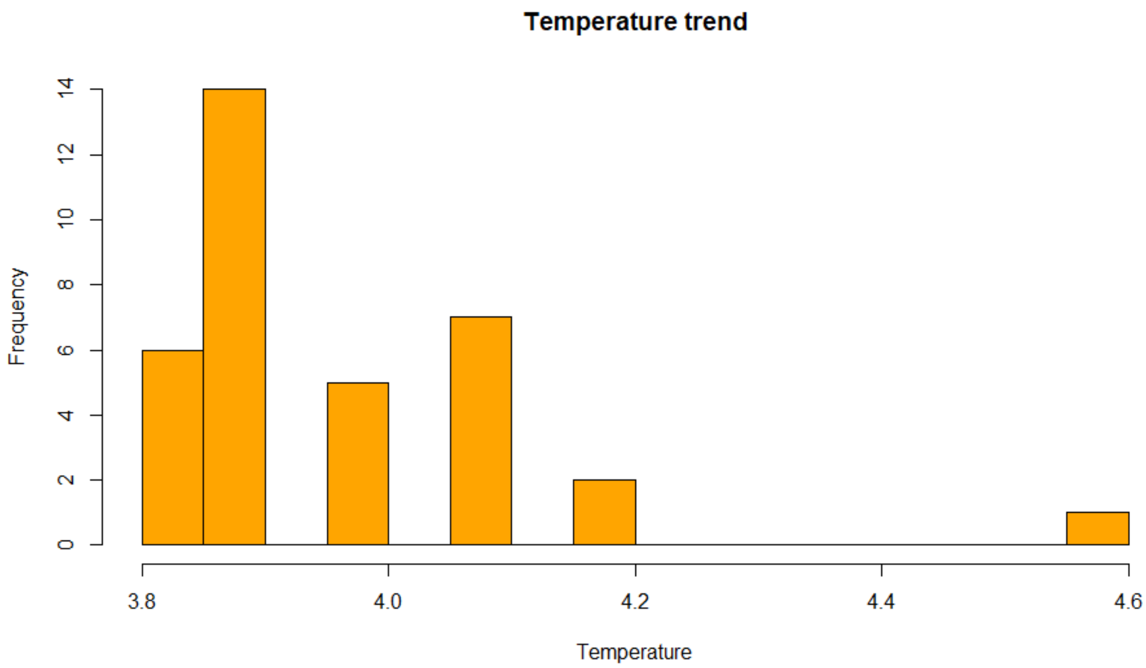
Histogram of the temperature variations from the January 2016 dataset .

Name of the dataset used : “Cold_Storage_Temp_Data.csv”



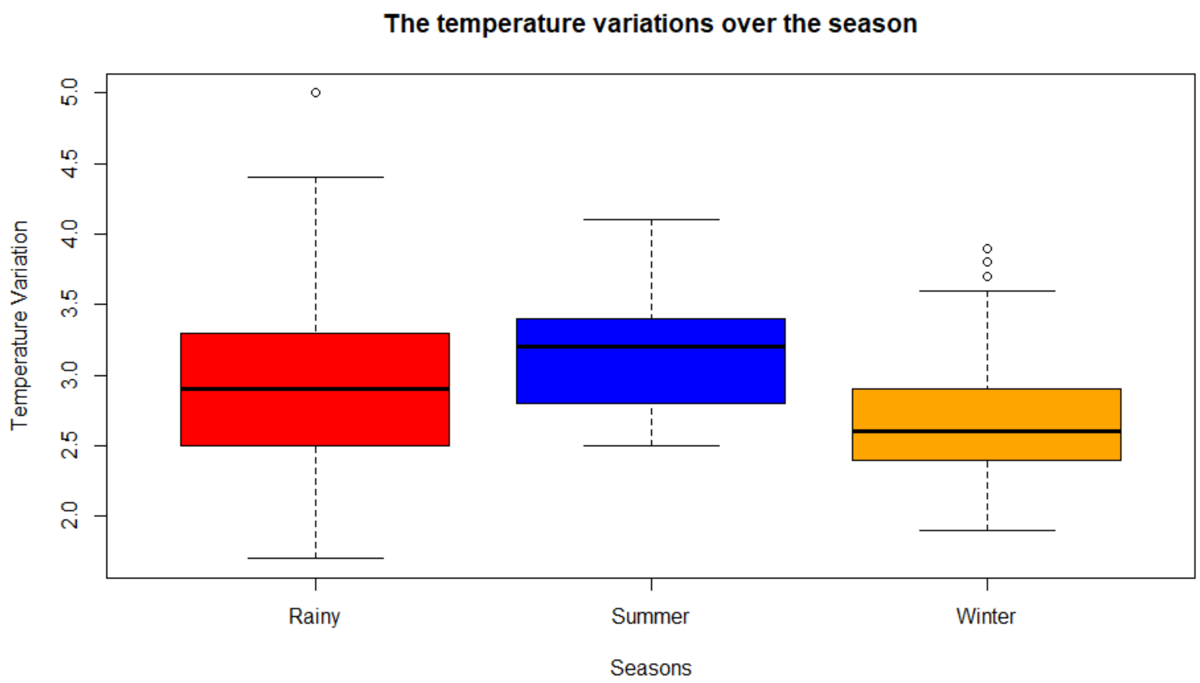
Histogram of the temperature variations for the March 2018 dataset.

Name of the dataset used : “Cold_Storage_Mar2018.csv”

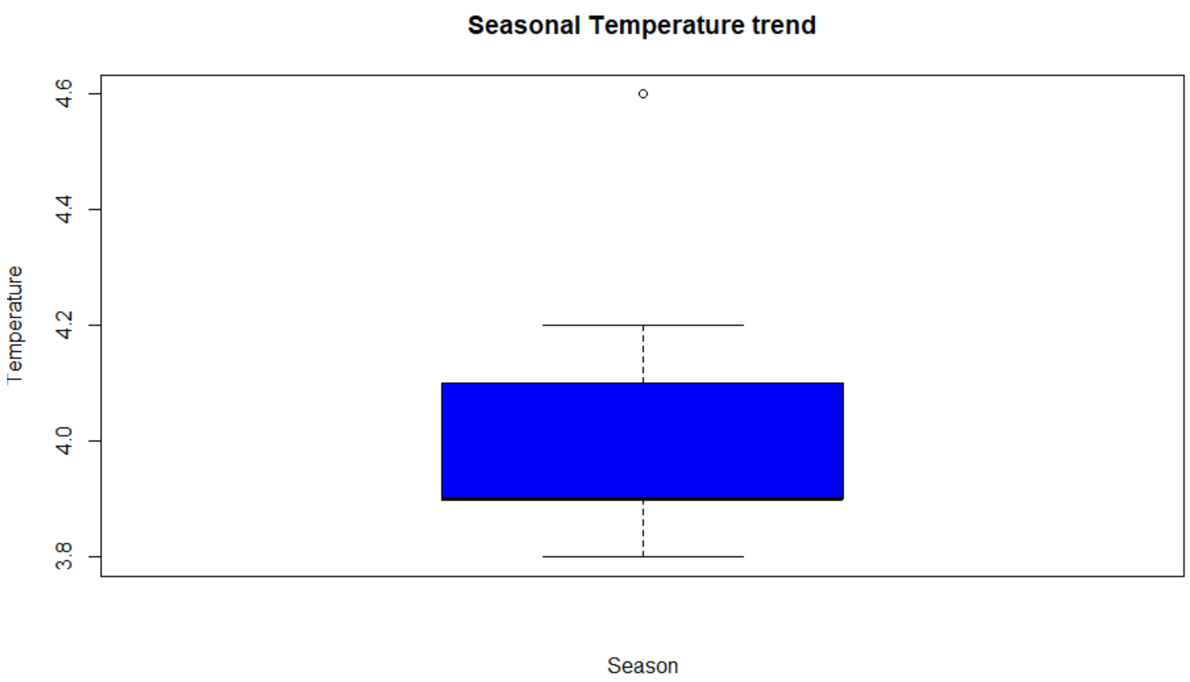


Both the temperature variations for the graphs state that neither of the dataset are positively or negatively skewed.

Season wise boxplot for the temperature variations :-



Boxplot for the March 2018 temperature dataset



3.5 Hypothesis Testing:

For Hypothesis Testing we have two options to go forward with and they are:-

1.Null Hypothesis (H_0)

2.Alternative Hypothesis (H_a)

Let's discuss what actually they mean.

Null Hypothesis: It is a hypothesis that says there is no statistical significance between the two variables. The null hypothesis is formulated such that the rejection of the null hypothesis proves the alternative hypothesis is true.

Alternative Hypothesis: It is one that states there is a statistically significant relationship between two variables. The alternative hypothesis is the hypothesis used in hypothesis testing that is contrary to the null hypothesis.

3.6 Observations:

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

Mean Winter Temperature = 2.7°C

Mean Summer Temperature = 3.15°C

Mean Rainy Temperature = 3.04°C

2. Overall mean for the full year

Overall mean = 2.96°C

3. Standard Deviation for the full year

SD for full year = 0.51

4. The probability of temperature having fallen below 2°C

Probability of Temp going below 2°C = 2.99%

5. The probability of temperature having gone above 4°C

Probability of Temp going above 4°C = 2.07%

6. The penalty for the AMC Company

Since probability of temperature going below 2°C is 2.99% which falls between 2.5% and 5% so there will be **10% of AMC fee**.

7. Hypothesis Testing - z-test

We have a single variable of temperature based on which we will perform the z-test and t-test hypothesis for the March 2018 dataset.
We will either reject or accept the null hypothesis in both cases

$\alpha = 0.1$, so Confidence level = $1 - \alpha = 0.9$

Data used as sample is = Cold_Storage_Mar2018.csv

Test performed = One sample z-test

Null Hypothesis (H_0): $\mu = 3.9$
Alternative Hypothesis (H_a): $\mu > 3.9$

```
        One Sample z-test

data:  cold_Storage_march_data$Temperature
z = 147.25, n = 35.00000, Std. Dev. = 0.15967, Std. Dev. of the sample mean = 0.02699,
p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 3.929891 4.018680
sample estimates:
mean of cold_Storage_march_data$Temperature
                                   3.974286
```

8. Hypothesis Testing - t-test

Data used as sample is = Cold_Storage_Mar2018.csv
Test performed = One sample t-test

alpha = 0.1, so Confidence level = 1 - alpha = 0.9

Null Hypothesis (H_0): $\mu = 3.9$
Alternative Hypothesis (H_a): $\mu > 3.9$

```
        One Sample t-test

data:  cold_Storage_march_data$Temperature
t = 147.25, df = 34, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 3.928648 4.019923
sample estimates:
mean of x
 3.974286
```

9. Inference after both the Tests

Both the tests (z-test and t-test) have **Rejected the Null Hypothesis** and **Accepted the Alternative Hypothesis** that temperature Mean value is above 3.9°C, so there are chances that the temperature in the cold storage can go beyond 3.9°C which is max limit and hence there is a need to take some action in the cold storage to correct this.

1. Conclusions

- For the yearly data of 2016
  - It is observed that as the season changes the mean temperature of the cold storage varies, i.e. more in summers and lesser in winters
  - Overall mean temperature is near the Minimum 2°C value and varies with a SD of 0.51

- There is a 2.99% probability that the temperature will go below 2°C and 2.07% probability that it will go above 4°C
- Since probability of temperature going below 2°C is 2.99% so AMC will be fined with 10% of their fee as per the agreement (10% penalty for 2.5% - 5% & 25% above 5%)

- For the March data of 2018
  - After getting complaints from consumers when concerns were raised, there was necessity to if the quality of the storage items is affected due to Cold storage temperature or it is from the procurement side
  - To address this z-test and t-test were performed on the data of 35 days to check if storage temperature is exceeding 3.9°C
  - Both the tests proved that the issue is in the Cold Storage itself and the temperature of storage is going beyond 3.9°C which is affecting the food items and there is need to take corrective measures in the facility

2. Appendix A – Source Code

Read the Datasets

```
cold_storage_data = read.csv("Cold_Storage_Temp_Data.csv", header = TRUE)
cold_march = read.csv("Cold_Storage_Mar2018.csv" , header = TRUE)
```

Import Packages

```
library(readr)

library(ggplot2)

library(TeachingDemos)
```

Summary of Cold_Storage_Temp_Data.csv

```
summary(cold_storage_data)

##     Season        Month          Date        Temperature
##  Rainy :122   Aug    : 31   Min.   : 1.00   Min.   :1.700
##  Summer:120   Dec    : 31   1st Qu.: 8.00   1st Qu.:2.500
##  Winter:123   Jan    : 31   Median :16.00   Median :2.900
##               Jul    : 31   Mean   :15.72   Mean   :2.963
##               Mar    : 31   3rd Qu.:23.00   3rd Qu.:3.300
##               May    : 31   Max.   :31.00   Max.   :5.000
##               (Other):179

str(cold_storage_data)

## 'data.frame':    365 obs. of  4 variables:
##  $ Season     : Factor w/ 3 levels "Rainy","Summer",..: 3 3 3 3 3 3 3 3
3 3 ...
##  $ Month      : Factor w/ 12 levels "Apr","Aug","Dec",..: 5 5 5 5 5 5 5
5 5 5 ...
##  $ Date       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Temperature: num  2.4 2.3 2.4 2.8 2.5 2.4 2.8 2.3 2.4 2.8 ...
```

Summary of Cold_Storage_Mar2018.csv

```
summary(cold_march)
```

```
##     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
```

```
str(cold_march)
```

```
## 'data.frame':    35 obs. of  4 variables:
##  $ Season     : Factor w/ 1 level "Summer": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Month      : Factor w/ 2 levels "Feb","Mar": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Date       : int  11 12 13 14 15 16 17 18 19 20 ...
##  $ Temperature: num  4 3.9 3.9 4 3.8 4 4.1 4 3.8 3.9 ...
```

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

```
cold_winter = cold_storage_data[cold_storage_data$Season == "Winter",]

mean_cold_temp = round(mean(cold_winter$Temperature),2)

## [1] 2.7

cold_summer = cold_storage_data[cold_storage_data$Season == "Summer",]

mean_summer_temp = round(mean(cold_summer$Temperature),2)

## [1] 3.15

cold_rainy = cold_storage_data[cold_storage_data$Season == "Rainy",]

mean_rainy_temp = round(mean(cold_rainy$Temperature),2)

## [1] 3.04
```

Yearly Mean Temperature

```
mean_yearly_temp = round(mean(cold_storage_data$Temperature),2)
mean_yearly_temp

## [1] 2.96
```

Yearly SD

```
sd_yearly_temp = round(sd(cold_storage_data$Temperature),2)
sd_yearly_temp

## [1] 0.51
```

Probability for temp < 2deg

```
normalisation = pnorm(2, mean = 2.96, sd = 0.509)
normalisation

## [1] 0.02989406
```

Probability for temp > 4deg

```
Normalisation_2 = 1 - pnorm(4, mean = 2.96, sd = 0.509)
normalisation_2
```

```
## [1] 0.02071425
```

Hypothesis Testing - z-test

alpha = 0.1 , so Confidence level = 1 - alpha = 0.9

Null Hypothesis (H0): μ = 3.9

Alternative Hypothesis (Ha): μ > 3.9

```
z.test(cold_Storage_march_data$Temperature, sd = sd(cold_Storage_march_dat
a$Temperature), y = NULL, mean = 3.9, conf.level = 0.9)
```

```
##
##  One Sample z-test
##
## data:  cold_Storage_march_data$Temperature
## z = 147.25, n = 35.00000, Std. Dev. = 0.15967, Std. Dev. of the
## sample mean = 0.02699, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  3.929891 4.018680
## sample estimates:
## mean of cold_Storage_march_data$Temperature
##                                    3.974286
```

Hypothesis Testing - t-test

alpha = 0.1 , so Confidence level = 1 - alpha = 0.9

Null Hypothesis (H0): μ = 3.9

Alternative Hypothesis (Ha): μ > 3.9

```
t.test(cold_Storage_march_data$Temperature, y = NULL, mean = 3.9, conf.lev
el = 0.9)
```

```
##
##  One Sample t-test
##
## data:  cold_Storage_march_data$Temperature
## t = 147.25, df = 34, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  3.928648 4.019923
## sample estimates:
## mean of x
##  3.974286
```