Statistical Methods of Decision Making – Cold Storage

[The Problem 3](#_Toc18160383)

[Problem 1 3](#_Toc18160384)

[Problem 2 3](#_Toc18160385)

[Solution – Problem 1 4](#_Toc18160386)

[a) Find mean cold storage temperature for Summer, Winter and Rainy Season (4 marks) 6](#_Toc18160387)

[b) Find overall mean for the full year 7](#_Toc18160388)

[c) Find Standard Deviation for the full year 7](#_Toc18160389)

[d) Assume Normal distribution, what is the probability of temperature having fallen below 2 C? 7](#_Toc18160390)

[e) Assume Normal distribution, what is the probability of temperature having gone above 4 C? 8](#_Toc18160391)

[f) What will be the penalty for the AMC Company? 8](#_Toc18160392)

[Solution – Problem 2 9](#_Toc18160393)

[a) Which Hypothesis test shall be performed to check the if corrective action is needed at the cold storage plant? Justify your answer. 11](#_Toc18160394)

[b) State the Hypothesis, perform hypothesis test and determine p-value 12](#_Toc18160395)

[c) Give your inference 14](#_Toc18160396)

# **The Problem**

## **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 - 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 the it was statistically proven that 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 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”

1. Find mean cold storage temperature for Summer, Winter and Rainy Season
2. Find overall mean for the full year
3. Find Standard Deviation for the full year
4. Assume Normal distribution, what is the probability of temperature having fallen below 2 C?
5. Assume Normal distribution, what is the probability of temperature having gone above 4 C?
6. What will be the penalty for the AMC Company?

## **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 last 35 days’ temperatures. As a safety measure, the Supervisor decides to be vigilant to maintain the temperature 3.9 C or below.

Assume 3.9 C as upper acceptable value for mean temperature and at alpha = 0.1 do you feel that there is need for some corrective action in the Cold Storage Plant or is it that the problem is from 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 the if corrective action is needed at the cold storage plant? Justify your answer.
2. State the Hypothesis, perform hypothesis test and determine p-value
3. Give your inference

# **Solution – Problem 1**

We have the following assumptions before starting the study of the population

* The temperature has been collected for complete one yearwhen the plant maintenance was outsourced. So, we are assuming the **procedures has been stabilized**
* The temperature values recorded has **no influence on the outside temperature** as the outside temperature could vary over a period of 12 months.

**Descriptive Statistics:**

Let’s import the data into R environment and do preliminary analysis using summary & descriptive statistics.

# Setup Working Directory

setwd("c:/GL Class/Solution Preparation/SMDM")

getwd()

#upload your dataset

cold\_temp = read.csv("Cold\_Storage\_Temp\_Data.csv", header = TRUE)

|  |
| --- |
| >dim(cold\_temp)  [1] 3654  >str(cold\_temp)  '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 2.3 2.2 2.4 2.8 2.5 2.4 2.8 3 2.4 2.9 ...  >summary(cold\_temp)  Season Month Date Temperature  Rainy :122 Aug : 31 Min. : 1.00 Min. :1.700  Summer:120 Dec : 31 1st Qu.: 8.00 1st Qu.:2.700  Winter:123 Jan : 31 Median :16.00 Median :3.000  Jul : 31 Mean :15.72 Mean :3.002  Mar : 31 3rd Qu.:23.00 3rd Qu.:3.300  May : 31 Max. :31.00 Max. :4.500  (Other):179 |
|  |
|  |
|  |

**Conclusion**

* Dataset has records for 12 months and these months have been grouped and categorised in 3 groups as Rainy, Summer and Winter.
* Summary has the **minimum** temperature as **1** Degree Celsius and **maximum** temperature as **4.5** Degree Celsius.

**Graphical Analysis:**

Box plots and histograms

par(mfrow=c(1,2))

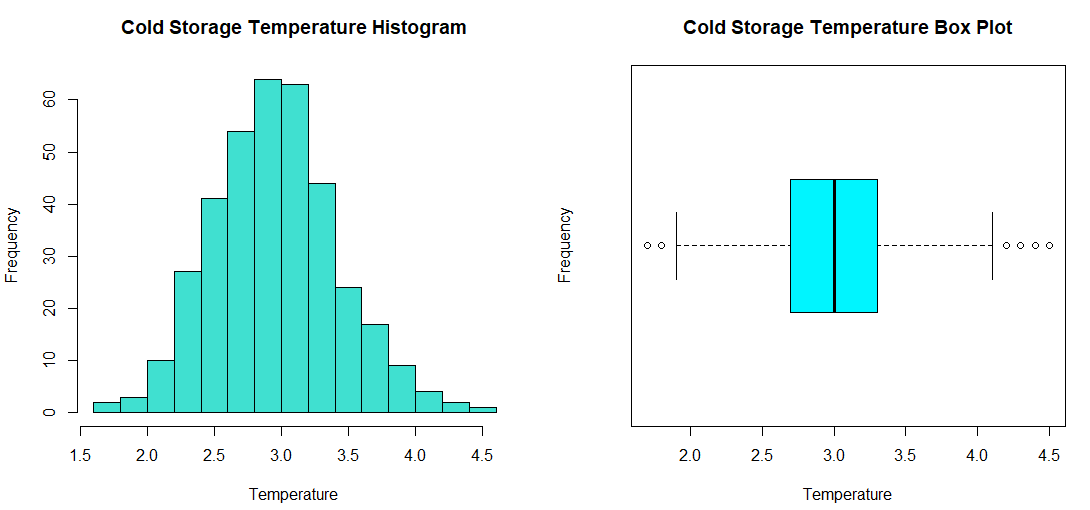
hist(Temperature,main='Cold Storage Temperature Histogram',xlab = "Temperature",

ylab = "Frequency",col = "turquoise")

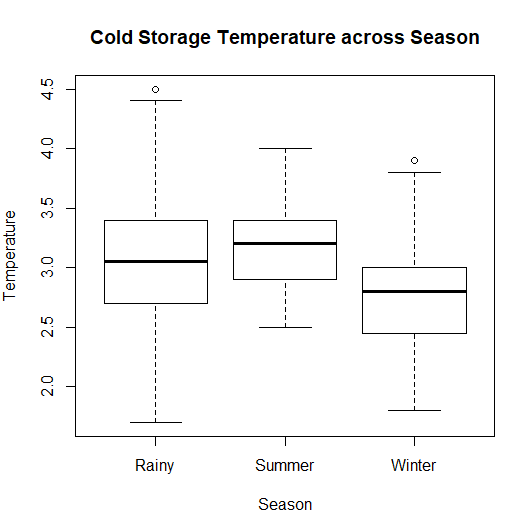
boxplot(Temperature,main='Cold Storage Temperature Box Plot',xlab = "Temperature",

ylab = "Frequency",col = "turquoise1",horizontal = TRUE)

dev.off()



plot(Temperature ~ Season, main='Cold Storage Temperature across Season')



**Conclusion**

Following is the conclusion from the graphical analysis.

1. Data set follows a normal distribution as skewness is not observed.
2. There are possible outliers/extreme values in the temperature
3. There is difference in minimum, maximum, 1st Quartile, 3rd Quartile and Median temperature value when the same was analysed across 3 seasons.

With descriptive statistics it seems, the***temperature has not gone below 2 Degree and/or above 5 Degree***  However, inferential statistics need to be applied to draw any conclusion on the population.

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

There are multiple ways this problem can be solved. We will create three subsets for each season and then find the mean.

# Get Subset of Winter Data

winter\_temp <- cold\_temp[which(cold\_temp$Season=="Winter"),]

mean\_winter\_temp <- mean(winter\_temp$Temperature)

mean\_winter\_temp

**[1] 2.776423**

# Get Subset of Summer Data

summer\_temp <- cold\_temp[which(cold\_temp$Season=="Summer"),]

mean\_summer\_temp <- mean(summer\_temp$Temperature)

mean\_summer\_temp

**[1] 3.1475**

# Get Subset of Rainy Data

rainy\_temp <- cold\_temp[which(cold\_temp$Season=="Rainy"),]

mean\_rainy\_temp <- mean(rainy\_temp$Temperature)

mean\_rainy\_temp

**[1] 3.087705**

## **Find overall mean for the full year**

Overall mean is the mean value of all values in the dataset

##Mean Temp (Full Year)

mean\_full\_year\_temp <- mean(cold\_temp$Temperature)

mean\_full\_year\_temp

**[1] 3.002466**

## **Find Standard Deviation for the full year**

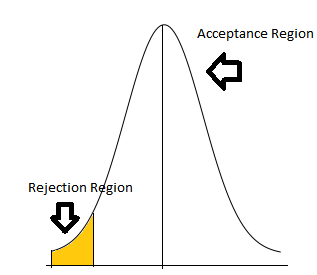
##Standard deviation for entire year

sd\_temp <- sd(cold\_temp$Temperature)

sd\_temp

**[1] 0.4658319**

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

As assumption is made that the distribution is Normal, we can use the function pnorm to find the probability. The probability needs to be calculated for values which have been fallen below 2 Degree C, this is a value which is shown in Yellow in the below diagram and hence lower tail is assumed to be true.

We have already calculated the mean and standard deviation value. Substituting these values, we get the following result

# calculate probability of temp <2

min\_temp <-2

#

prob\_less\_2 <- pnorm(q = min\_temp,

mean = mean\_full\_year\_temp,

sd = sd\_temp, lower.tail = TRUE )

prob\_less\_2

**[1] 0.01569906**

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

Now in this problem we have to calculate the value greater than 4 Degree C hence the calculations would be similar to (d) section of the problem with an only difference that lower.tail=FALSE.

# calculate probability of temp >4

max\_temp <-4

prob\_greater\_4 <- pnorm(q = max\_temp,

mean = mean\_full\_year\_temp,

sd = sd\_temp, lower.tail = FALSE)

prob\_greater\_4

**[1] 0.01612075**

## **What will be the penalty for the AMC Company?**

Let’s understand the penalty condition in terms of probability of temperature going outside the 2 - 4 C:

1. Above 2.5%and less than 5% then the penalty would be 10% of AMC (annual maintenance case).
2. In case it exceeded 5% then the penalty would be 25% of the AMC fee.

So let’s calculate the total probability for condition less than 2 Degree or greater than 4 Degree, i.e. sum total of probability calculated in (d) and (e) part of the problem

## Combined Probability for <2 and >4

Total\_Probability <- prob\_greater\_4+prob\_less\_2

Total\_Probability

**[1] 0.03181981**

Now use this calculated probability and check against the two conditions mentioned above to get the penalty

## Calculate Penatly

if(Total\_Probability>0.025 &&Total\_Probability <= 0.05) {

penalty <- '10%'

}else if(Total\_Probability >0.05){

penalty <- '25%'

}else

penalty <- '0%'

penalty

**[1] "10%"**

# **Solution – Problem 2**

We have the following assumptions before starting the study of the sample

* The sample collection is **random** and not **biased** in nature
* The sample is **not affected by outside temperature**

**Descriptive Statistics:**

Let’s import the data into R environment and do preliminary analysis using summary & descriptive statistics.

#Upload data set

cold\_mar\_data = read.csv("Cold\_Storage\_Mar2018.csv", header = TRUE)

|  |
| --- |
| >dim(cold\_mar\_data)  [1] 354  >str(cold\_mar\_data)  '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 ...  >summary(cold\_mar\_data)  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 |
|  |
|  |
|  |

**Conclusion**

* Sample has 35 records, hence as per central limit theorem, sample could be assumed to be normally distributed
* Summary has the **minimum** temperature as **3.8**Degree Celsius and **maximum** temperature as **4.6** Degree Celsius.

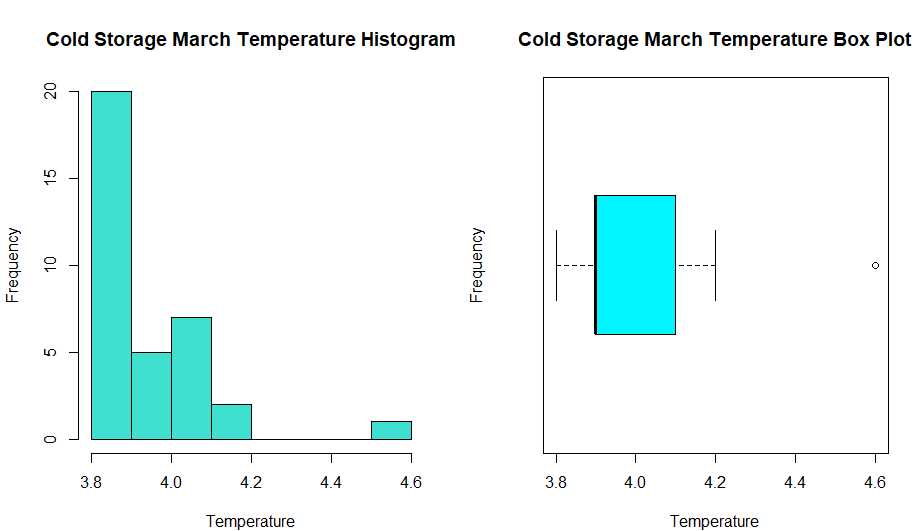
**Graphical Analysis:**

par(mfrow=c(1,2))

hist(Temperature,main='Cold Storage March Temperature Histogram',xlab = "Temperature", ylab = "Frequency",col = "turquoise")

boxplot(Temperature,main='Cold Storage March Temperature Box Plot',xlab = "Temperature", ylab = "Frequency",col = "turquoise1",horizontal = TRUE)

dev.off()



**Conclusion**

Following is the conclusion from the graphical analysis.

1. There are possible outliers/extreme values in the sample
2. It seems, the ***temperature has goneabove 3.9 Degree.*** However, inferential statistics need to be applied to draw any conclusion on the sample.

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

As the hypothesis testing requires to check if the men temperature of cold storage has gone above 3.9 Degree, hence two different type of tests could be performed

1. T-test
2. Z-test

However, for performing Z test we require the Population Standard Deviation. .Also Z test assumes that sample size is large enough and can be rightly compared with the population.

If we assume the Standard deviation calculated in Problem (1) to be the population values, Z test can be applied

Hence, we would perform both the test

As we are looking at only one sample, hence one sample test would be performed. Whether this would be a right tailed test or left tailed test, the same would be determined through alternate hypothesis.

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

Let’s start the Z test

**Hypothesis for Z- Testing**

**H0:µO<= 3.9**[Temperature is maintained below or equal to 3.9 Degree andhence no corrective action is required.]

**HA:µO> 3.9** [Temperature is above 3.9 Degree andhence corrective action is required.]

**Where**

**µO**is mean of the sample data set

As alternate hypothesis is with a greater sign, it is a **Right tailed** test,

As mentioned in part (a) of Problem (2), standard deviation of the complete dataset calculated in Problem (1) would be used as a Population standard deviation hence **population SD ( σ ) = 0.4658319**

#Z Score

z <- (mean - mu)/(sd\_temp/(sqrt(n)))

mean of the sample is

mean <- mean(cold\_mar\_data$Temperature)

mean

[1] 3.974286

Mu=3.9

n=35

Substituting these values in the above mentioned formula we get a Z score as

**0.9434308**

Given value of alpha =0.1 P value is

pValue = pnorm(z,lower.tail = FALSE)

pValue

**[1] 0.1727303**

As P-value is > given alpha, hence we **Fail to Reject the Null Hypothesis**, which indicates that temperature of the cold storage was maintained below 3.9 Degree and hence no corrective action is required.

Let’s perform the t test

**Hypothesis for Z- Testing**

**H0:µO<= 3.9**[Temperature is maintained below or equal to 3.9 Degree andhence no corrective action is required.]

**HA:µO> 3.9** [Temperature is above 3.9 Degree andhence corrective action is required.]

**Where**

**µO**is mean of the sample data set

As alternate hypothesis is with a greater sign, it is a **Right tailed** test,

With alpha=0.1, confidence interval is .9, mu=3.9

t.test(cold\_mar\_data$Temperature, mu

= mu,

alternative = "greater",

conf.level = conf)

One Sample t-test

data: cold\_mar\_data$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

P value is < alpha and hence **Null Hypothesis is rejected** which indicates that temperature of the cold storage has gone above 3.9 and hence corrective action is required.

## **Give your inference**

A test statistic is a numerical summary of the data that is compared to what would be expected under the null hypothesis. Test statistics can take on many forms such as the z-tests (usually used for large datasets) or t-tests (usually used when datasets are small).

Even though we have adequate sample size to perform Z test, we don't know the standard deviation of the population. So, Z test is not appropriate here.

Hence the output obtained through t-test would be appropriate and hence corrective action is recommended for cold storage.

|  |
| --- |
|  |