# Mini Project - Cold Storage

Name : Numer P

# Table of Contents

## 1. Project Objective

The main objective of the report is to explore the Cold Storage Dataset ("Cold_Storage_Temp_Data") and ("Cold_Storage_Mar2018.csv") in R and generate insights about the data set. This exploration report will consist of the following,

- Importing dataset in R
- Understanding the structure of Dataset
- Graphical exploration
- Descriptive Statistics

## 2. Assumptions

In cold Storage operations, the penalty clauses will be higher in the Summer when compared with the winter and rainy season. This may affect the AMC Fee in handling the milk products. If the temperature 2-4 C is maintained for the products, then the AMC will extend the contract for next year.

The client complaints from the end customers are computed with alpha = 0.1 and mu = 3.9 which relates the one tailed hypothesis and will reject the H0 for the 35 days data. The Hypothesis will turn the P-Value will leads to level of significance when the Z is less than 5%.

## 3. Exploratory Data Analysis – Step by Step Approach

1. Environment Set up and Data Import
2. Variable Identification
3. Univariate Analysis
4. Bivariate Analysis
5. Variable Transformation
6. Feature Exploration

## 3.1 Environment Setup and Data Import

### 3.1.1 Install necessary packages and Import Libraries

This section is used to install packages and invoke the associated libraries. Having all packages at the same places increase code readability.

### 3.1.2 Setup Working Directory

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.

Please refer Appendix A for source code.

### 3.1.3 Import and read the dataset

The given dataset is in .csv format. Hence the command 'read.csv' is used for import the dataset.

Please refer Appendix A for source code.

## 3.2 Variable Identification

- setwd() used for setup working directory to export data and files from the folder or location in PC.
- getwd() used to identify the location was correctly entered or not.
- Library function is used to load the installed packages like ggplot2, dplyr, rpivotTable, readxl, readr and mice.
- read_csv function is used to load the csv files in the path.
- sd function is used to find the standard deviation of the cold storage dataset.
- attach function is used to reduce the reusability of variable name to enter each time.
- str function is used to check the category variables formats.
- summary function is used to get the summarised value like length, class and basic statistics values with quartile ranges.
- rpivotTable function is used to get the individual mean values in cold storage dataset in table format.
- pnorm function is used to find the probability of temperature between 2 C and 4C.
- dim function is used to find the total observations and variables.
- by function is used to sort out the category variables.

### 3.2.1 Variable Identification – Inferences

*#getwd()*

[1] "/Users/numerp/Documents/PGP-BABI/Project -1 (Cold Storage Case Study) (SMDM)"

*#library(mice)*

(mice)takes out cbind and rbind the variables of two different datasets.

*#library(readr)*

(readr) helps in reading the rectangular datasets while the datasets are in table format.

*#library(readxl)*

(readxl) helps in reading the files in excel formats.

*#library(dplyr)*

(dplyr) helps in filter the datasets and intersect the datasets.

*#library(ggplot2)*

(ggplot2) helps is visualise the datasets in boxplot, histogram and graphical representation.

*#read_csv*

csv file was imported from the path and shows the variables

*#str*

str shows the variables along with class of the data. It shows some samples to understand the data.

*#class*

class function describes the full file in data.frame format. As the files includes category in season variables, it shows the values as character format.

*#attach*

Variable is attached to reduce the reusability in following syntax.

*#dim*

It shows the number observations and the variables associated in the file.

*#summary*

It produces the results as summarised format for each variable.

*#sd*

Standard deviation is calculated for the dataset.

*#pnorm*

It shows the probability of the temperature falling between the ranges of 2 C and 4 C.

*#by*

It is used to sort out the categories of season from the dataset and its function is used to find out the summary of each category variables.

*#Z.test in excel*

The Z test which is used to find the one tailed hypothesis testing.

*#Norm.Dist in excel*

Normal Distribution can be find in excel using norm.dist function.

## 3.3 Univariate Analysis

Univariate analysis is the analysis of data of one variable at time and it involves whether the datasets are descriptive or inferential statistics.

- **Mean cold storage temperature for summer, winter and rainy season**

  Mean is the central values for the dataset. The average mean value for the season are in the below table.

| Season | Summer | Winter | Rainy |
|---|---|---|---|
| Mean | 3.153 | 2.701 | 3.039 |

- **Overall mean for the full year**

  The overall mean for the cold storage temperature maintained is 2.963

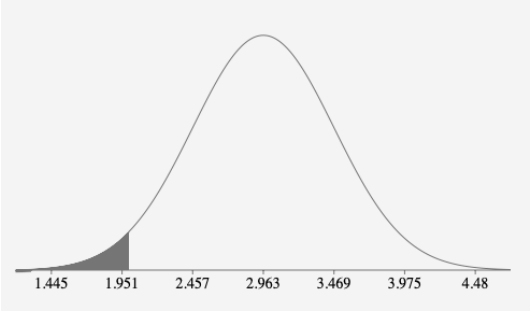
| Temperature – Mean |
|---|
| 2.963 |

- **Standard deviation for the full year**

  Standard deviation is the measure of variation of variables in the datasets.

| Temperature – Standard Deviation |
|---|
| 0.508589 |

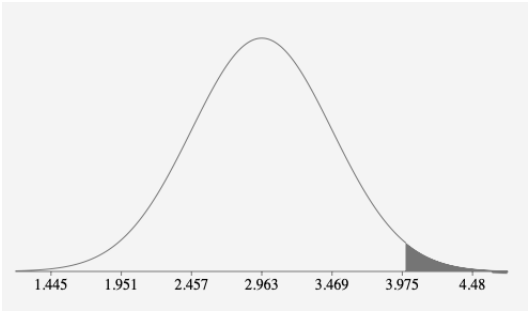
❖ **Probability of temperature having fallen below 2 C**




The normal distribution shows that the probability of temperature fallen below 2 C is 3% which helps the AMC company to protect store the items as not higher possibility of frozen state. Hence, the temperature below 2 C is not affecting the items and also it decreases the penalty for the storage company.

| **Probability of temperature having fallen below 2 C** |
|---|
| Probability value - 0.02914744 |

❖ **Probability of temperature having gone above 4 C**



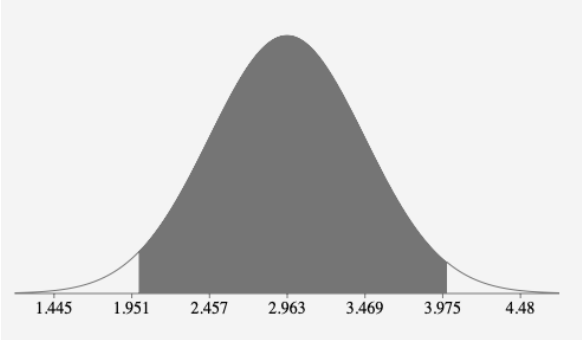

The normal distribution for the cold storage temperature having gone above is 97% and the value shows that items are not maintained in the ideal state i.e. between 2C-4C. In this case, the company may increase the chance of paying penalty for not maintaining the temperature in ideal state.

| **Probability of temperature having gone above 4 C** |
|---|
| Probability value - 0.0207821 |

### ❖ Penalty for AMC Company

The penalty for the AMC company will not be affected. The normal distribution for the cold storage temperature range between 2 C – 4 C is 25%. This shows that the AMC Company penalty will be 10%. But the probability for the temperature above 4 C shows 0.2% which is to bared by the company to decrease the penalty.

In this case, the penalty will be 10% for annually as per the normal distribution shows the probability between 2 C and 4 C.




### ❖ Hypothesis test performed to check the corrective action is needed at the cold storage plant

The statistical hypothesis is a statement that used to assume or predict the variable for the datasets.

Null Hypothesis: Mu<=3.9

Alternative Hypothesis: Mu>3.9

1. A null hypothesis is the status quo. It is formulated when the rejection leads to desired conclusion which is arrived in the alternative hypothesis.
2. The null hypothesis performs the different types of errors namely Type I error and Type II error
3. Alpha is the probability of rejecting the null hypothesis when it is true.
4. Beta is the probability of accepting the null hypothesis when it is false.
5. (1-Alpha) is the probability of accepting the null hypothesis when it is true.
6. Where, alpha is the level of Significance of the test.

One tailed Z Test is used to find the corrective action needed in the storage plant.

## One tailed Z Test

The term one tailed test signifies the all or Z – Values that would cause the rejection or accepting the null hypothesis.

Hence, the hypothesis testing follows as, mu equal to 3.9 and mu not equal to 3.9.

- Interpretation

The null hypothesis is formulated for the samples of 35 days which is an average of 3.974 and the standard deviation of 0.159. The alternative hypothesis is less than the computed average mean as per the storage company. The assumed temperature shows that the products stored in the range of 3.9 C which is below or maintained equally. This datasets are to tested using one tailed hypothesis which is Z test to perform the normal distribution of the temperature. The Z test returns the p-value which is mostly predicted to be less than alpha. If the p value is less than the alpha then the null hypothesis should be rejected and the p value greater than alpha, it states that fail to reject null hypothesis.

The alpha is 0.1 which is 10% of the normal distribution and it should be used in accepting the error or rejecting the null hypothesis. If the p value is less than 0.1, then products stored in the storage container is met with problem and the company should replace it with other one. The rejection helps in finding the procurement team have maintained the products at good stage and while storage is showing the damaged status.

If the p value is higher than 0.1, then the storage container is maintained in good state and the company have to check the procurement team to replace the transport system to get the products to storage. This shows the one tailed Z test is importance in brings the p value to promotes the structure of the values.

The level of significance in P value is trends report that it is less than 0.1 then the storage container is definitely having problem. This significance level is taken when the p value is less than the 0.1. Hence the significance level shows the real problem in storing the products.

❖ Findings

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

1. X bar for the temperature in 3.974286
2. Standard deviation for the temperature is 0.159674
3. The one tail z test value is 2.752359
4. The significance level of p value is 0.00471119

The significance level of p value is less than the alpha value 0.1. Then the findings is failed to reject the H0. It shows that the storage is not having problem and the corrective measure have to be taken in the procurement side to avoid the customer complaints.

❖ **Inferences**

The significance p value is less than alpha and it concludes to fails the reject the Null Hypothesis.

The gold storage for the dairy products is maintained in good condition while stored in the company and it reflects the procurement side is having problem. The cold storage company should maintains their transportation time from the source and makes the dairy products not to melt and makes it not smelly and taste sour. The Safety measure in procurement side should be maintained with AC contained truck to transport the diary goods and to regularly change the dairy container. Procurement team should collect their goods in major sources from nearer warehouse to store the dairy products in the cold storage.

The samples shows the values are maintained in cold storage have to be corrected in each category however, the procurement team transport the dairy products in short time it is important in cold storage plant to maintains the temperature below 3.9 C will makes the customers to be attracted for the dairy products of the company.

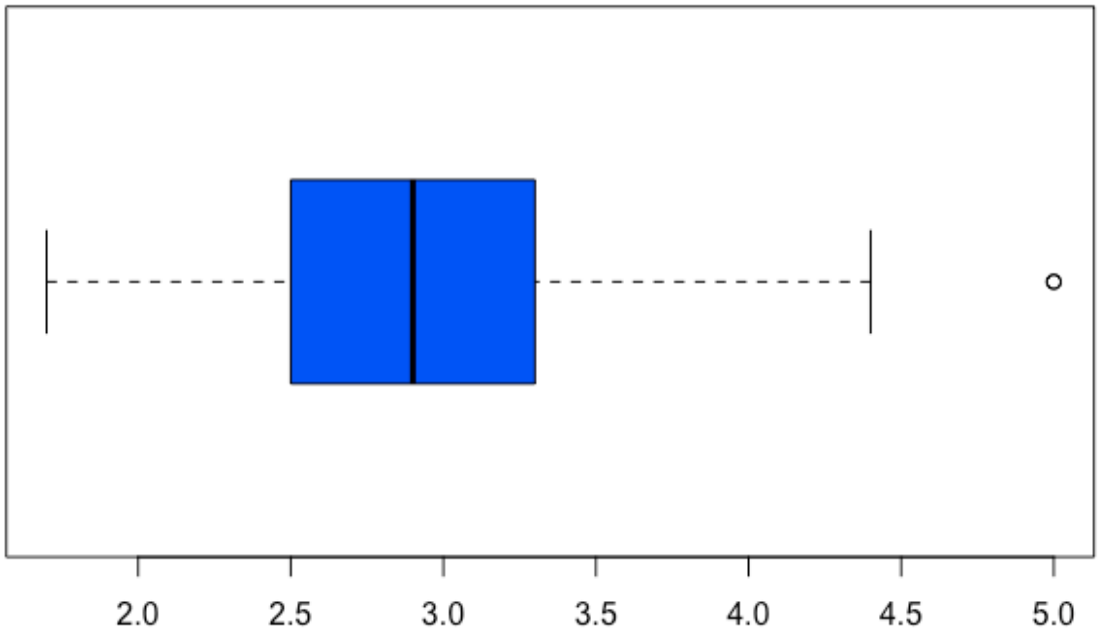
## 3.4 Bivariate Analysis

Bivariate Analysis is used to analyse the two variables and find the relationship between them. This analysis will help in identifying the association and strength of the variables. The analysis is used find the correlations, distributions and scatter plot.

As the cold storage case study is mostly focus on one dependent variable (Temperature), hence the bivariate analysis is not possible in finding the values for the datasets.

## 3.5 Outlier Identification

The outlier identification for the temperature is found above 4 as the maximum value for the temperature dataset.




## 3.6 Variable Transformation/ Feature creation

The variables are suitable to analyse the and interpret the findings to help the temperature differences in the old storage.

It will be better for the different dataset on comparing the temperature maintained in procurement side will helps to find more understanding in maintaining the values for the AMC Company.

## 4. Conclusion

The AMC company is signed for the contract of maintaining the dairy products in the cold storage between the ideal temperature of 2 C – 4 C. There is penalty for not maintaining the products as per the contract which is 10% for annually and if 25% for the temperature exceeds than 4 C. The AMC Company is keeping the diary products where they get it from the contractor to provide the better quality in storing the dairy products. This includes the procurement side of transporting the goods to the storage plant. The descriptive statistics shows that the company is maintaining the dairy products in good condition. While analyse the datasets, it shows that the cold storage temperature is fallen between 2 C – 4 C, this proves the overall cold storage is good at 95%. This will lead the company to penalty of 10% annually as it said in the contractor.

The supervisor findings on the customer complaints in March and doubted the problem occurs in procurement or storage plant. When analyse the march temperature it shows the products are in good condition while stored in storage plant as well as stored in the procurement side. The data which is provided for the storage plant again proves the same kind of good conditioned products and can predict the problem may occur in procurement side. The procurement side can find the problem to control the temperature while transporting the materials for the storage plant.

It is concluded that the customer complaints in march will makes penalty to the storage company, as they have not met the requirements in transportation (procurement). Hence penalty mentioned 10% can be taken over form AMC company and the corrective measures taken in procurement team for the next transportation will makes the contractor to sign the storage contractor for next year. This helps in understanding the company positive side and the negative in transporting the dairy products. This customer complaints will be redirected with the penalty amount and can be replaced with the free maintenance charges for a month and will continue business to next year. The cold storage inferences are showing this interpretation is better in making the contractor and the business leading person to makes more profitable from the storage plant.

## 5. Appendix

**PROBLEM 1 – R SOURCE CODE**

```
setwd("/Users/numerp/documents/PGP-BABI/Project -1 (Cold Storage Case Study) (SMDM)")
getwd()
library(mice)
library(readr)
library(readxl)
library(ggplot2)
library(dplyr)
read_csv("Cold_Storage_Temp_Data.csv")
temperature=read_csv("Cold_Storage_Temp_Data.csv")
str(temperature)
class(temperature)
View(temperature)
summary(temperature)
sd(temperature$Temperature,na.rm=F)
head(temperature,5)
tail(temperature,10)
attach(temperature)
dim(temperature)
names(temperature)
temeperature=as.data.frame(temperature)
class(temeperature)
by(temeperature,INDICES = Season,FUN = summary)
a=by(temeperature,INDICES = Season,FUN = summary)
a
pnorm(2,mean=2.962739726,sd=0.508589031,lower.tail = T,log.p = F)
z=pnorm(2,mean=2.963,sd=0.508589,lower.tail = T,log.p = F)
y=pnorm(4,mean=2.963,sd=0.508589,lower.tail = T,log.p = F)
z
y
boxplot(temeperature$Temperature)
boxplot(temeperature$Temperature,horizontal = T)
boxplot(temperature$Temperature,horizontal = T,col="blue")
```

**PROBLEM 2 – R SOURCE CODE**

```
setwd("/Users/numerp/documents/PGP-BABI/Project -1 (Cold Storage Case Study) (SMDM)")
library(mice)
library(readr)
library(readxl)
library(ggplot2)
library(dplyr)
hypothesis=read_csv("Cold_Storage_Mar2018.csv")
View(hypothesis)
summary(hypothesis)
str(hypothesis)
hypothesis=as.data.frame(hypothesis)
class(hypothesis)
#Hypothesis Testing: H0:Mu=3.9 H1=Mu not equal to 3.9
hypothesis
attach(hypothesis)
xbar=mean(Temperature)
xbar
s=sd(Temperature)
s
mu=3.9
n=35
tstat=(xbar-mu)/(s/(n^0.5))
tstat
pvalue=pt(tstat,34)
pvalue
pvalue=(1-pt(tstat,34))
pvalue
```