COLD STORAGE CASE STUDY

Estimation and Hypothesis Testing Project

Chandrima Nandy
26th January 2020

Abstract

To assess the current temperature regulation system of Cold Storage Plant and recommend any scope of improvement.

Table of Contents

Background

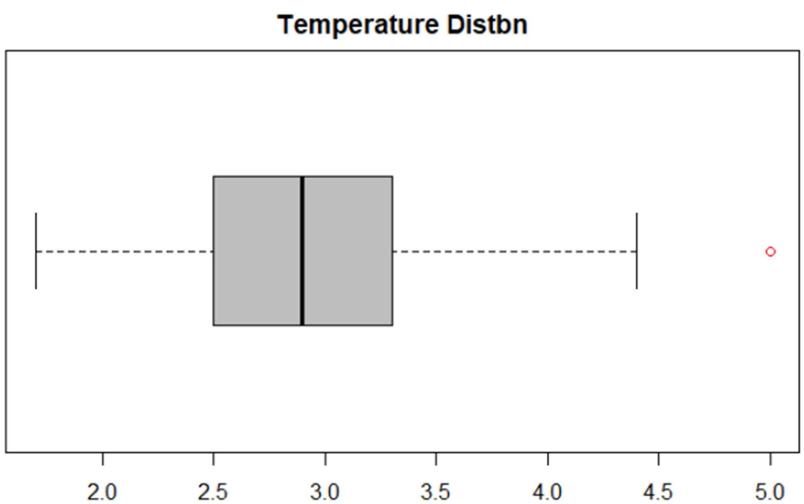
Cold Storage started its operations in Jan 2016. They are in the business of storing Pasteurized Fresh Whole or Skimmed Milk, Sweet Cream, Flavoured Milk Drinks. To ensure that there is no change of texture, body appearance, separation of fats the optimal temperature to be maintained is between 2 - 4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

Problem Statement – 1

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 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".

Q1) Find mean cold storage temperature for Summer, Winter and Rainy Season

Dataset exploration of Cold_Storage_Temp_Data.csv revealed the presence of outliers in the distribution of the variable Temperature; specifically, two outliers with the value of 5 are present in the data. Therefore, the mean has been computed twice as given in the below table:



a) Once by keeping the given data unchanged as given by MeanBySeason

b) After adjusting the outliers by capping their value at Q3+1.5IQR as given by MeanBySeason_Adjusted

Season <fctr>	MeanBySeason <dbl>	MeanBySeason_Adjusted <dbl>
Rainy	3.039344	3.031148
Summer	3.153333	3.153333
Winter	2.700813	2.700813

3 rows

We see that the presence of outliers has only a slight influence in changing the mean. Only for Rainy season there is a minor variation in mean temperature while mean temperature for Summer and Winter has remained the same before and after changing the outlier value.

Q2) Find overall mean for the full year

As before the overall mean for the full year has been computed twice -

a) Keeping the given Temperature data unchanged, the mean was 2.96274

b) Capping the Temperature outlier value at Q3+1.5IQR, the mean was found to be 2.96

We find a slight variation in the mean value because of the presence of outliers.

Q3) Find Standard Deviation for the full year

As before the overall standard deviation (s.d.) for the full year has been computed twice -

a) Keeping the given Temperature data unchanged, the s.d. was 0.508589

b) Capping the Temperature outlier value at Q3+1.5IQR, the s.d. was found to be 0.4988338

Again, we find a slight variation in the s.d. value because of the presence of outliers.

Q4) Assume Normal distribution, what is the probability of temperature having fallen below 2 C?

Since we found that the presence of outliers hasn't had a significant influence in varying the mean and standard deviation, therefore the original Temperature data has been considered in computing the probability value.

Given the Normal distribution assumption and with the prior calculated mean and standard deviation values, we have used the 'pnorm' function to compute the probability of the temperature falling below 2C which came out to be 0.02918142 or 2.91%.

The underlying logic followed is:

If x is a normally distributed random variable, with mean = μ and standard deviation = σ,

then P(x < xmax) = pnorm(xmax, mean = μ, sd = σ, lower.tail=TRUE)

Q5) Assume Normal distribution, what is the probability of temperature having gone above 4 C?

Again, following the same reasoning as mentioned above, the original Temperature data has been considered in computing the probability value.

Given the Normal distribution assumption and with the prior calculated mean and standard deviation values, we have used the 'pnorm' function to compute the probability of the temperature rising above 4C which came out to be 0.02070079 or 2.07%.

The underlying logic followed is:

If x is a normally distributed random variable, with mean = μ and standard deviation = σ,

then P(x > xmin) = pnorm(xmin, mean = μ, sd = σ, lower.tail=FALSE)

Q6) What will be the penalty for the AMC Company?

The probability of the temperature falling below 2C and the temperature rising above 4C are two mutually exclusive events i.e. they can't occur at the same time. Therefore, the probability of the temperature going outside the range 2 - 4 C (implying the temperature to be either less than 2C or greater than 4C) is given by the sum of individual probability of occurrence of these two events,

i.e. 0.02918142 + 0.02070079 = 0.04988221 or 4.99%

The underlying logic followed is: P(A U B) = P(A) + P(B)

Since the probability came out to be greater than 2.5% but less than 5%, therefore the penalty would be 10% of AMC (annual maintenance case).

Problem statement - 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".

Q1) Which Hypothesis test shall be performed to check that if corrective action is needed at the cold storage plant? Justify your answer.

As per the given Temperature data in Cold_Storage_Mar2018.csv file, we find the following:

Sample size (n)=35, hypothesized mean value (M)=3.9,

sample mean ($\overline{x}$)=3.974286, sample standard deviation (s)=0.159674,

degrees of freedom = n-1 = 34, level of significance(α)=0.1

As the sample size >=30, therefore we can assume the sample to follow normal distribution (using the Central Limit Theorem).

Since we do not have the population standard deviation value, therefore we need to conduct a one sample upper tailed T test. A one-sample t-test is used to test whether a population mean is significantly different from some hypothesized value.

Q2) State the Hypothesis, perform hypothesis test and determine p-value

The null hypothesis is given as :

Ho: Population mean is less than or equal to 3.9 or μ<=3.9

The alternative hypothesis is given as :

Ha : Population mean is greater than 3.9 or μ>3.9

The t-statistic is given by :

$t = (\bar{x} - M)/ (s/ \sqrt{n})$

The hypothesis testing has been performed twice:

a) Once by taking the Temperature field unchanged (i.e. keeping the outliers as they are).

   The result as provided below, gives the t-statistic to be 2.7524 and the p-value to be 0.004711.

```
```{r}
t.test(Mar2018_df$Temperature,mu=3.9, alternative = "greater")
```

	One Sample t-test

data: Mar2018_df$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
```

b) Again by adjusting the outliers present in Temperature field by capping their value at 0.95 quantile. The result as provided below, gives the t-statistic to be 3.002 and the p-value to be 0.0025.

```
```{r}
t.test(Mar2018_df$TemperatureAdjusted,mu=3.9, alternative = "greater")
```

	One Sample t-test

data: Mar2018_df$TemperatureAdjusted
t = 3.002, df = 34, p-value = 0.0025
alternative hypothesis: true mean is greater than 3.9
95 percent confidence interval:
 3.927452   Inf
sample estimates:
mean of x
 3.962857
```

As we see, the t-test has been performed using t.test function. Since the alternative hypothesis suggests an upper tailed test, alternative="greater" is mentioned in the function.

Q3) Give your inference.

In both the t-tests mentioned above we found the p-value to be less than the level of significance of 0.1 which leads us to accept the alternative hypothesis that the population mean exceeds the acceptable temperature level of 3.9C. Therefore there is need for some corrective action in the Cold Storage Plant for temperature regulation.