# Project Objective

Cold Storage started operations in January 2016, they specialise in storing pasteurized fresh whole or skimmed milk, sweet cream, flavoured milk drinks, but to ensure that there is no change of texture, body appearance, separation of fats, the company agreed for an optimal temperature between 2 - 4 c to be maintained. However, during the first year of business they outsourced the plant maintenance work to a professional company ‘Coldplay Services’. As a result of the optimal temperature of (2 – 4 c) that is to be maintained, Cold storage decided to place stiff penalty clauses.

It was agreed that if it statistically proven that the probability of the temperature going outside the range of (2 – 4 c) during the contract timeline of contract one year was above 2.5% and less than 5% (2.5% > 5%), Coldplay Services would forfeit 10% of the Annual maintenance contract (AMC). Nevertheless, if it was more than 5% then the penalty would be 25% of the AMC for Coldplay Services.

Having collected the data set of annual temperature recordings for 2016 the report would analyse the average temperature and conduct a statistical analysis to see if Coldplay Services should be given full AMC and to continue services with Cold storage.

# Assumptions

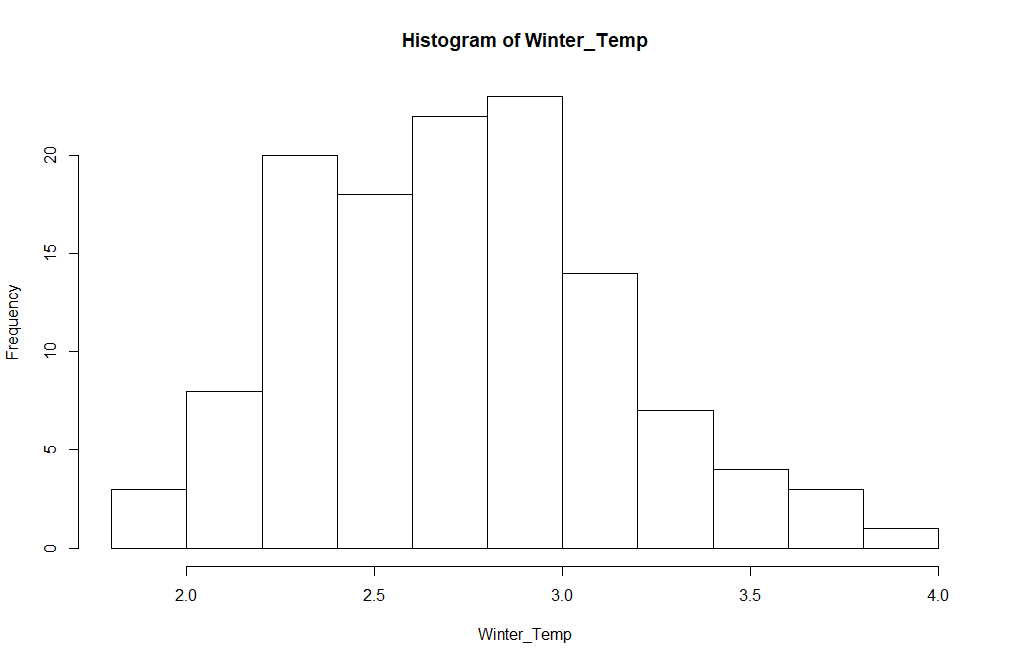
Giving the dataset for the year we made the following assumptions

* That the Annual Maintenance Cost for outsourcing is $2.5 million
* That the data recorded is a normal distribution
* That the mean at least one is slightly different from the others
* That Season could affect the increase in temperature degrees
* That Coldplay Services would be awarded penalty charges as it is impossible to have not gone above 2.5%
* That the temperature recorded according to months or date could cause outliers in the data

# Analysing the means

This section analyses the average cold storage temperature recorded during each season.

## Winter Season

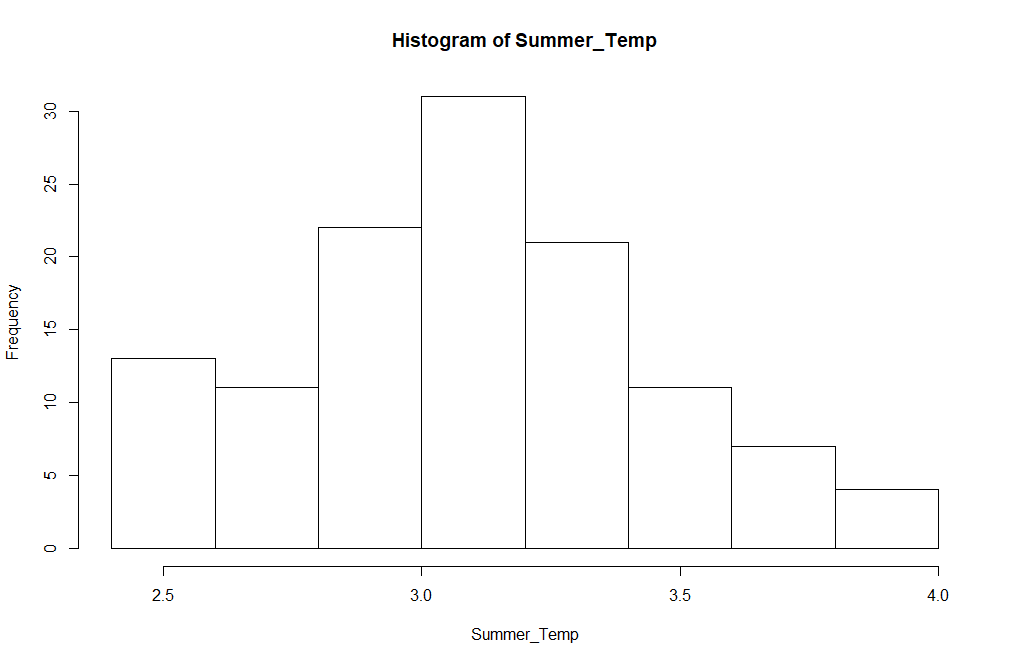


The winter season from the dataset is between October to January 2012, from the Histogram, we can recordings of a high frequency in winter temperature between 2.3 – 3 C , which could mean that the winter season might have had an effect on the storage temperature, storage items could be affected if the data points where increased at high frequency between 3.5 to 4C or had more frequent recording less than 2.3C. The average winter temperature is 2.776 which is slightly more skewed towards the outlier values. The average winter temperature may be quite inaccurate because of the outlier the data and the skewness is slightly skewed to the right showing a positive direction, the median is slightly larger than the mean and this could be as result of the outlier as outliers shift the mean in an unrealistic way. But the mean temperature is still significant

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.800 2.450 2.800 2.776 3.000 3.900

## Summer Temperature



The temperature of Summer falls within the months February - May 2012, the average temperature during the summer is 3.147 C there was an increase in summer temperature compared to the mean recorded during winter season. Again, this cold be as a result that during the summer the level of cool is highly affected so to maintain the normal cool there needed to be an increase. The summer temperature ranges from a minimum of 2.5 C to a maximum of 4 C. However the mean, median, mode are also slightly ,they thereabout fall within the same range making the data an normal distribution, we are choosing to ignore the minimum and maximum values because of the nature of the histogram, this could be that 4 C was recorded because of the item where to froze and needed to reduced so as to keep the goods from getting maybe separation of fats, change in texture etc. the mean temperature is still very significant and accurate, identifying the bell curve as a normal distribution eliminates all outliers within the data. the median is slightly larger than the mean and this could be as result of the 4 C recorded during the summer.

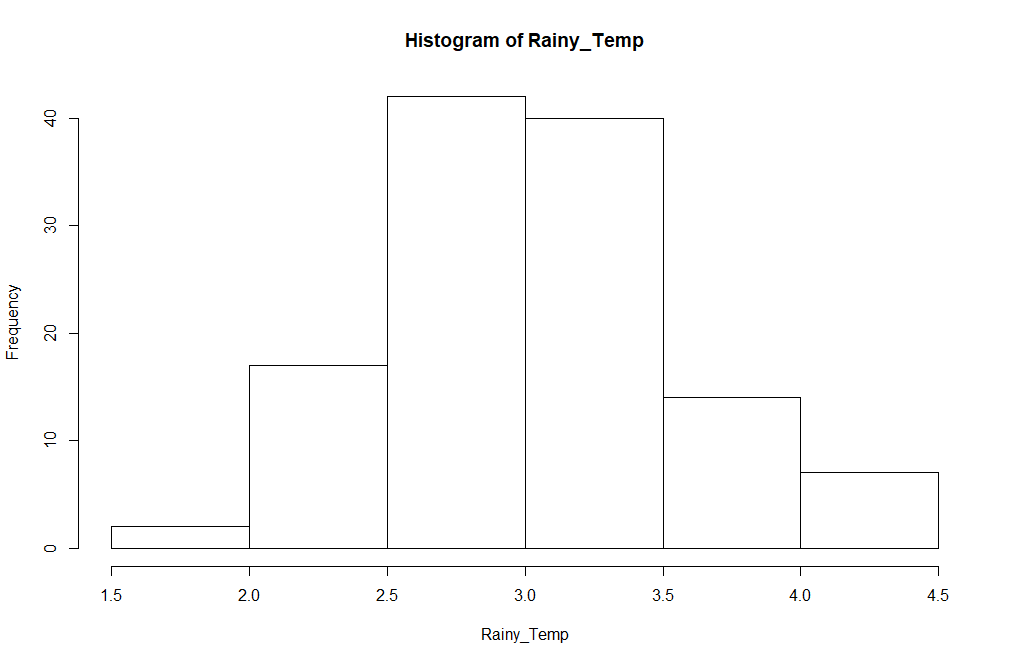
Mode= 3.2

summary(Summer\_Temp)

Min. 1st Qu. Median Mean 3rd Qu. Max.

* 1. 2.900 3.200 3.147 3.400 4.000

## Rainy Temperature



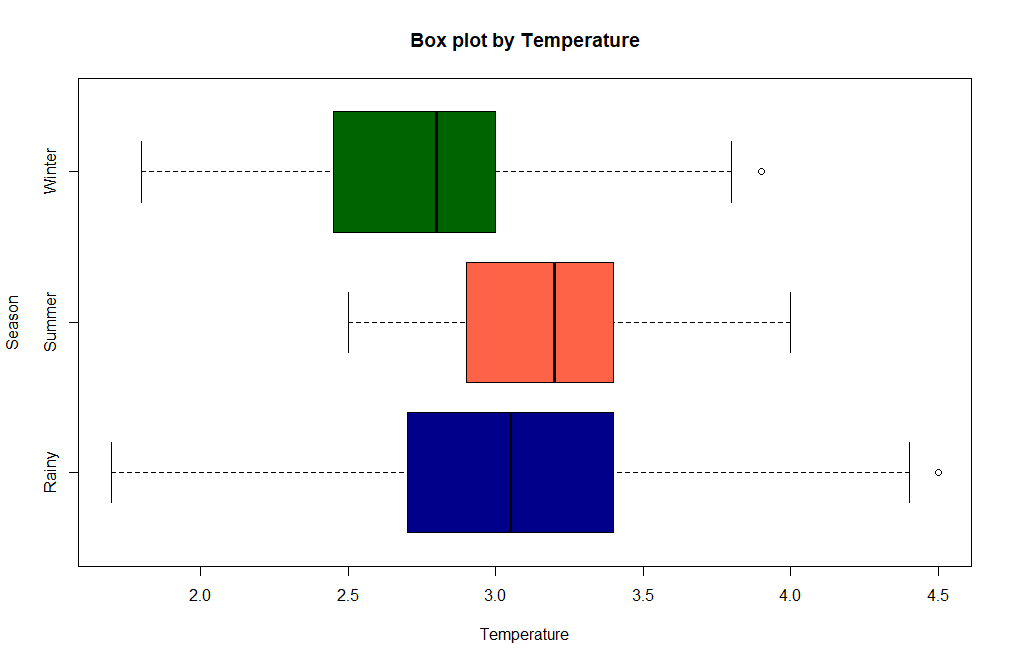
The winter season from the dataset is between June to September 2012, The dataset records data from a minimum 0f 1.7 to a maximum of 4.5 during this season. From the Histogram, we see that the data is slightly skewed to the left, negative, showing the presence of an outlier. The average winter temperature is 3.088 which is slightly more skewed towards the outlier values. The average winter temperature may be quite inaccurate because of the outlier the data and the skewness are slightly skewed to the right showing a positive direction, the median is slightly larger than the mean and this could be as result of the outlier as outliers shift the mean in an unrealistic way. But the mean temperature is still significant

Mode = 2.9

summary(Rainy\_Temp)

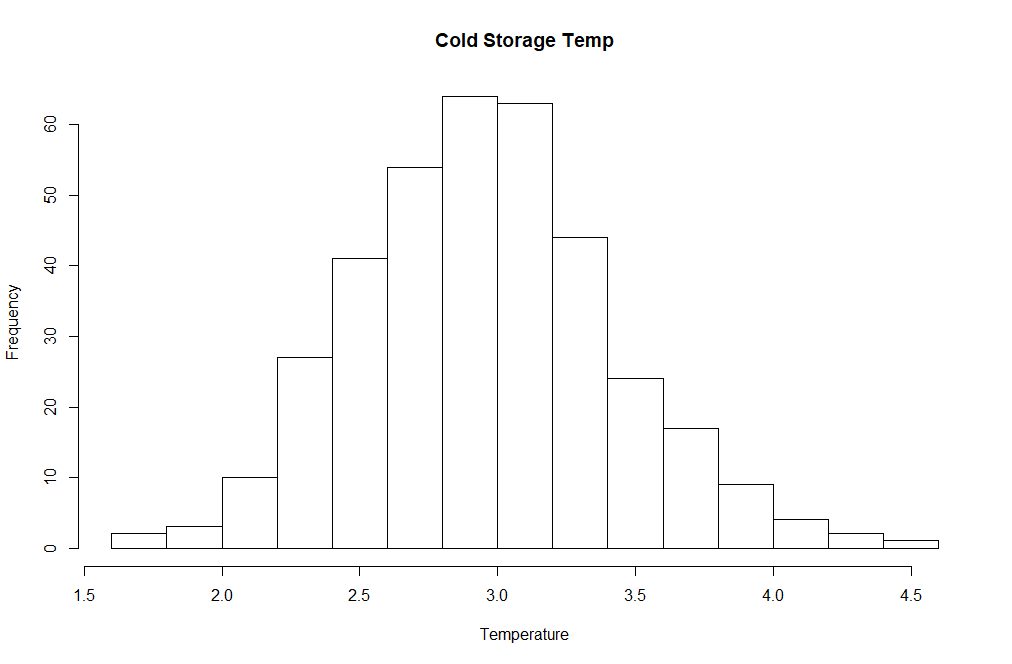
Min. 1st Qu. Median Mean 3rd Qu. Max.

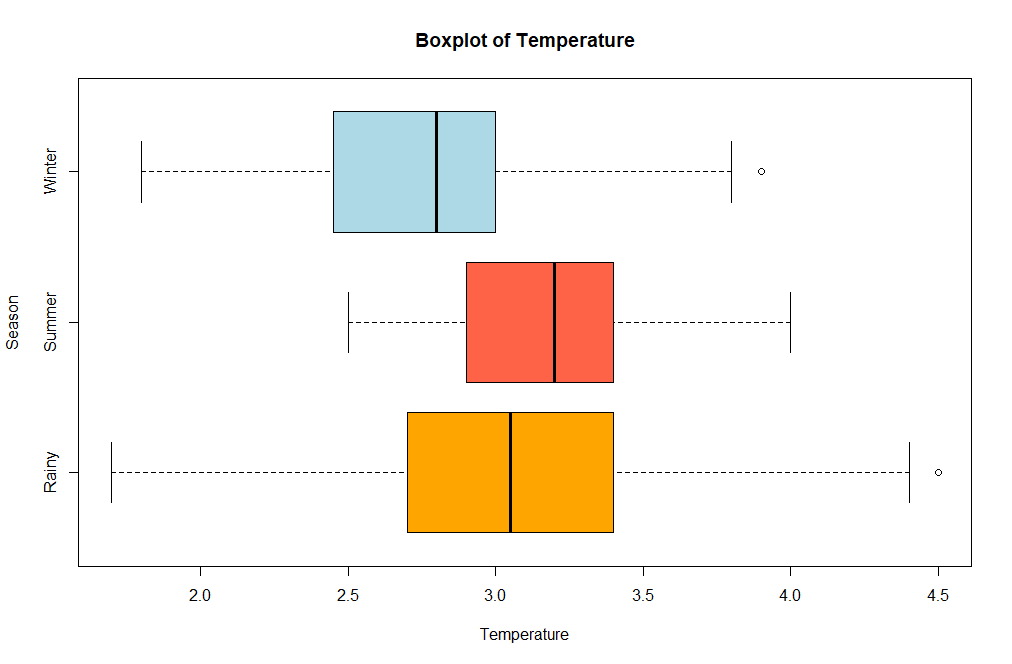
1.700 2.700 3.050 3.088 3.400 4.500



Winter IQR = 0.55

# Average Annual Temperature





Standard deviation

Variance

# Statistical Analyses

## Normal Distribution

### Finding the Probability that the temperature fell below 2C

Using the average temperature value of 3 and a standard deviation of 0.466 and an X value of 2

# Anova – Analysis of variance

*HO: All pairs of group means are equal against*

*HA: At least one group mean is different from the rest*

𝐻0: 𝜇1 = 𝜇2 = 𝜇3

𝐻A: At least one temperature level is different from the rest.

Before one-way ANOVA procedure is applied to the data, visual comparison is recommended. Moreover, the normality and equality of variance assumptions need to be checked.

# Question 2

# Project Objective

# T- test

A sample T- test .

Paired sample t-test which compares means from the same group at different times.

One sample t-test which tests the mean of a single group against a known mean.

it lets you know if those differences (measured in means/averages) could have happened by chance.

The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups.

A large t-score tells you that the groups are different.

A small t-score tells you that the groups are similar

Every t-value has a p-value to go with it. A p-value is the probability that the results from your sample data occurred by chance. P-values are from 0% to 100%. They are usually written as a decimal. For example, a p value of 5% is 0.05. Low p-values are good; They indicate your data did not occur by chance. For example, a p-value of .01 means there is only a 1% probability that the results from an experiment happened by chance. In most cases, a p-value of 0.05 (5%) is accepted to mean the data is valid.

Sample Size: 30, No. of samples: 2  Measurement repeated on the same sample after 4 months.  Five Point Summary and Standard Deviation on both the samples.

Describe the five per cent significance test you would apply to these data to determine whether new scheme has significantly raised outputs?

 The level of significance (Alpha ) = 0.05.  The sample size , N = 30 which is sufficiently large for a Zstat Test.  But since the population standard deviation (Sigma) is unknown, we have to use a Tstat test.  Degree of Freedom: Since the sample is the same for both Sampling tests, we have N-1 degrees of freedom : 29  Since the sole purpose of the test is to check whether the New scheme is successful compared to old scheme, we would prefer a Right tailed T Test.

# Calculating the hypothesis

*Ho =*

*Ha =*

Observations: o Both the samples seems to be normally distributed.

o Mean and Median Values are not much different

Conclusion: o In this scenario, the p value is 0.065 which is greater than the 0.05. Hence accept the null hypothesis that the new scheme did not raise the output significantly.

2