Storage Temperature Report

Cold Storage Co.

Submitted by:
Mohanned T. Gomaa
8 November 2019

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 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 (3 marks)*
2. *Find overall mean for the full year (3 marks)*
3. *Find Standard Deviation for the full year (3 marks)*
4. *Assume Normal distribution, what is the probability of temperature having fallen below 2 C? (6 marks)*
5. *Assume Normal distribution, what is the probability of temperature having gone above 4 C? (6 marks)*
6. *What will be the penalty for the AMC Company? (7 marks)*
7. *Perform a one-way ANOVA test to determine if there is a significant difference in Cold Storage temperature between rainy, summer and winter seasons and comment on the findings. (9 marks)*

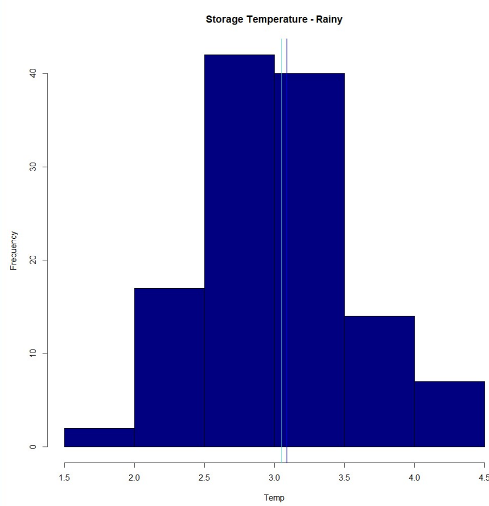
Problem 1 – Question 1

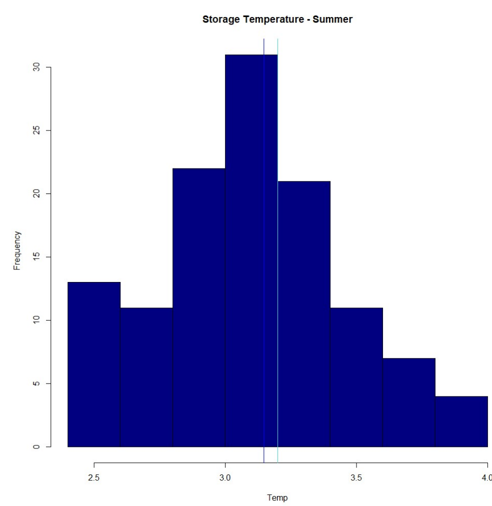
Average cold storage temperature per season:

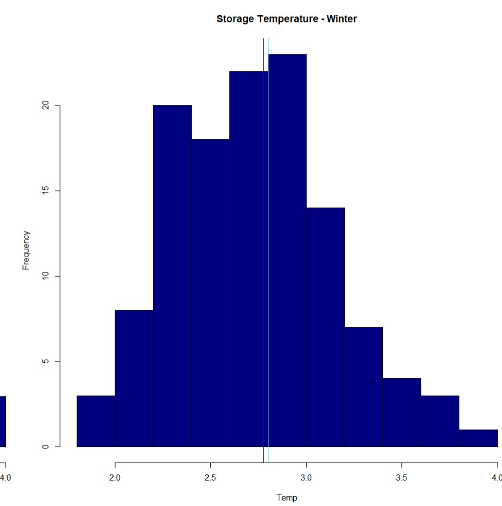
3.087705
Rainy Season

3.1475
Summer Season

2.776423
Winter Season



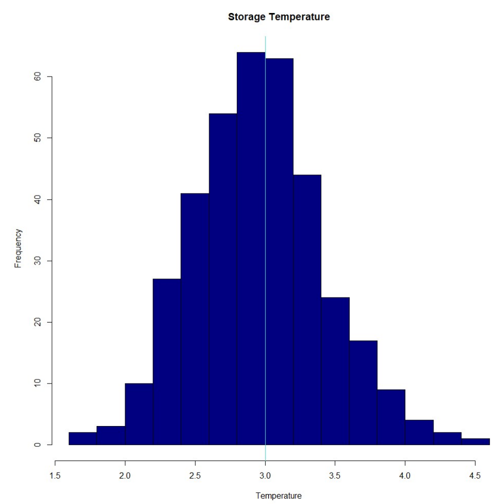




Problem 1 – Question 2

Average cold storage temperature for the whole year:

3.002
Whole Year

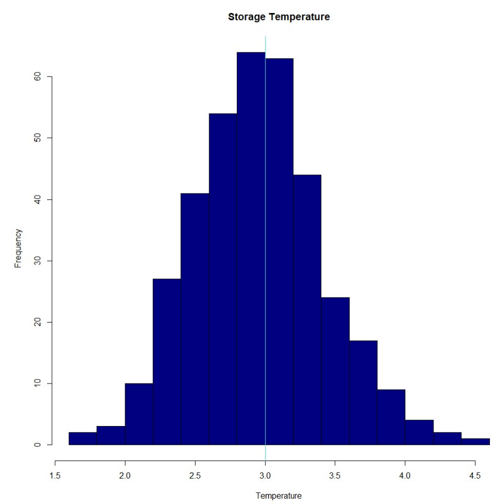


Temperature data seems to normally distributed, no visible signs of skewness.

Problem 1 – Question 3

Standard Deviation cold storage temperature for the whole year:

0.4658
Whole Year

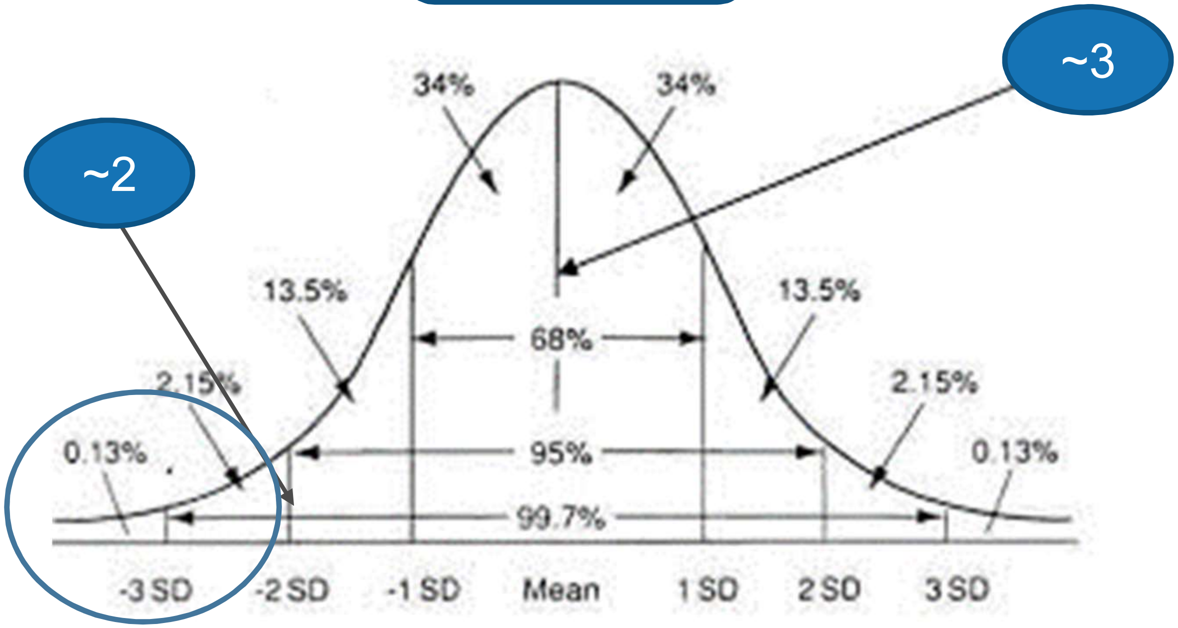


Temperature data seems vary from the mean by ~ 0.47 C.

Problem 1 – Question 4

The probability of temperature have fallen below 2C:

1.57%
P:Temp>2C

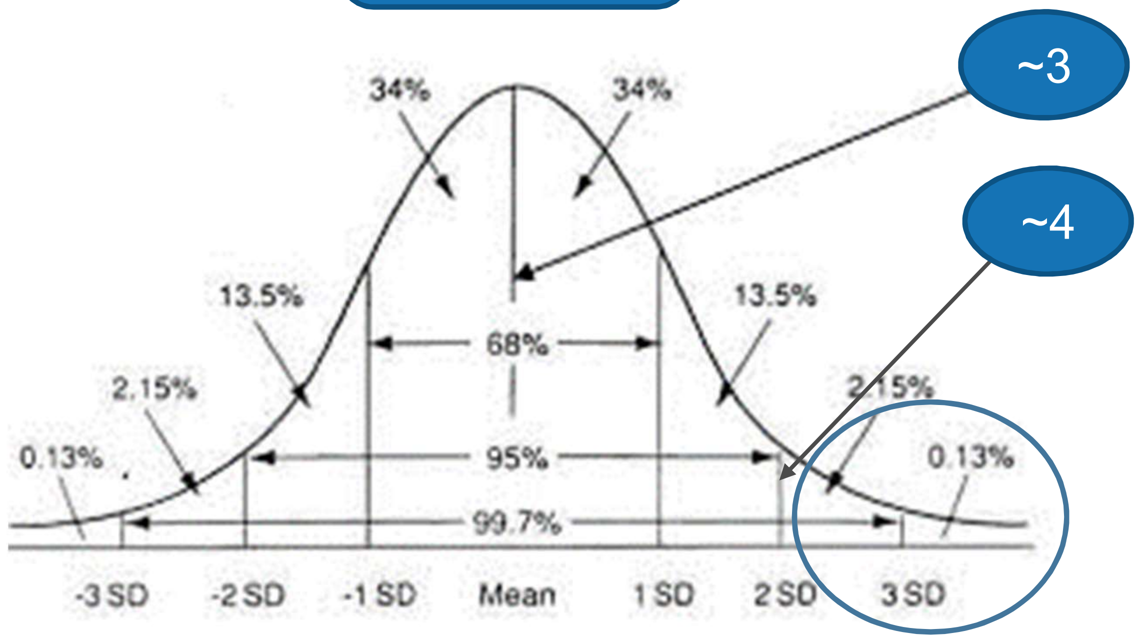


Assuming that temperate data is normally distributed, the above graph explains the distribution of temperature data. Where the median is ~ 3C and the STD is ~ 0.47C, so this mean that 2C is nearly -2SDs from the mean. That being said the probability is 1.57%.

Problem 1 – Question 5

The probability of temperature goes above 4C:

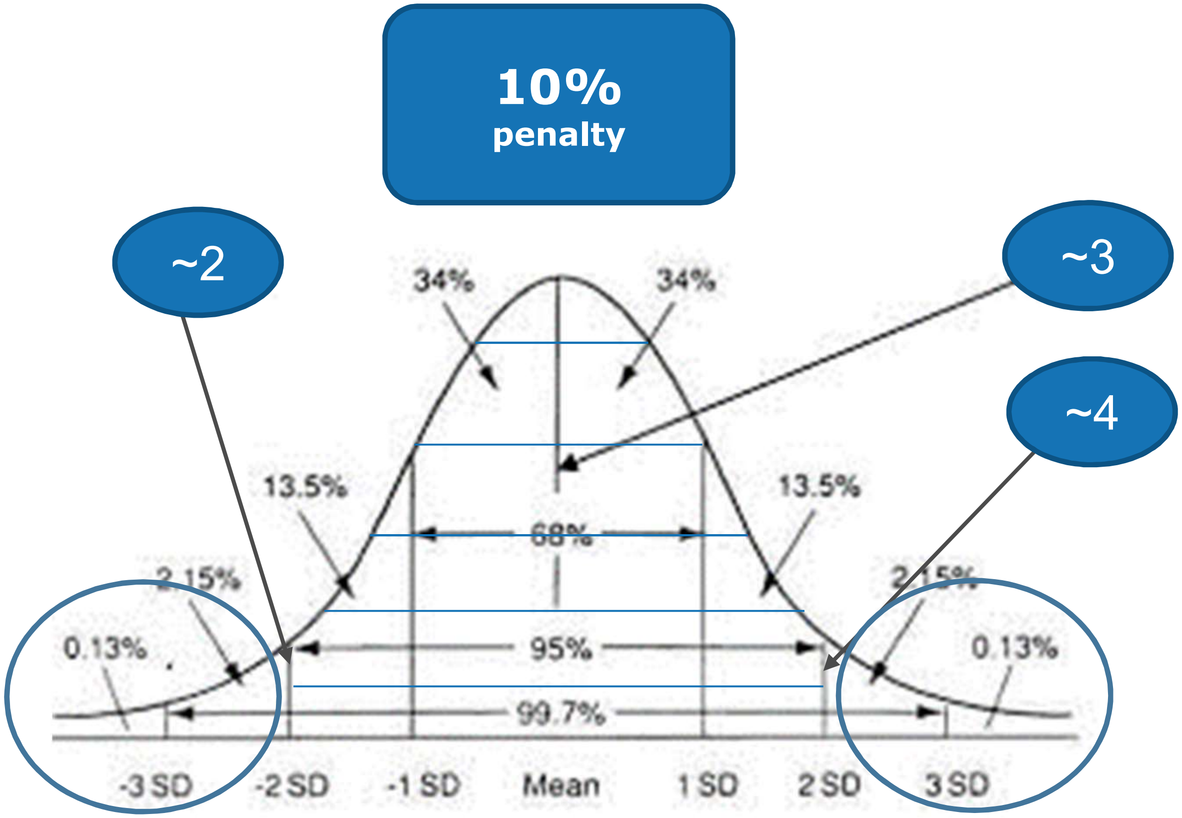
1.61%
P:Temp<4C



Assuming that temperate data is normally distributed, the above graph explains the distribution of temperature data. Where the median is ~ 3C and the STD is ~ 0.47C, so this means that 24 is nearly 2SDs from the mean. That being said the probability is 1.6%.

Problem 1 – Question 6

AMC Company penalty :

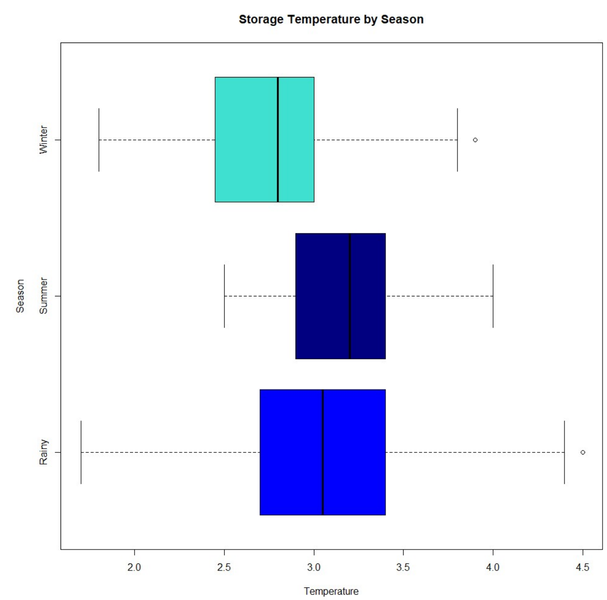


Assuming that temperate data is normally distributed, the above graph explains the distribution of temperature data. Where the range between 2C to 4C covers ~95% of the distribution, then the probability for temperature falling out of this range cant exceed 5%, so the penalty will be 10% of the AMC contract fees.

Problem 1 – Question7

Are all season temperature mean equal :

H0: All seasons temperature means are equal
Ha: H0 is not true



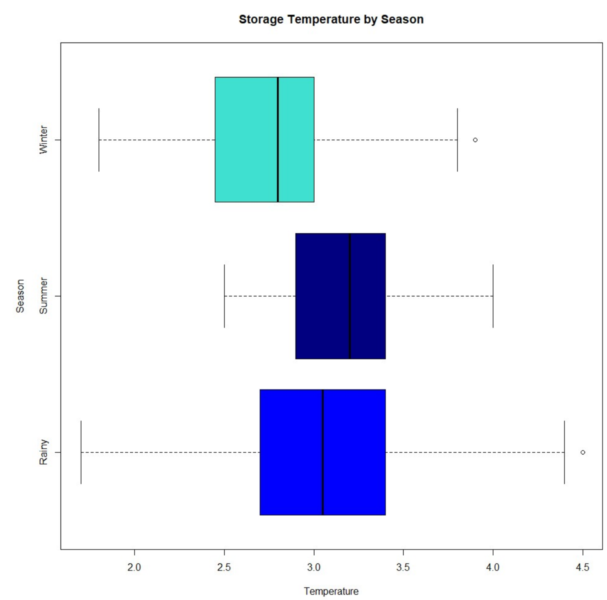
The boxplot shows the temperature data distribution by season and it seems to be of a similar distribution, with small expectation of winter.

Problem 1 – Question7

Are all season temperature mean equal :

0.05044
P-Value Normality

0.0003951
P-Value Homogeneity of variance



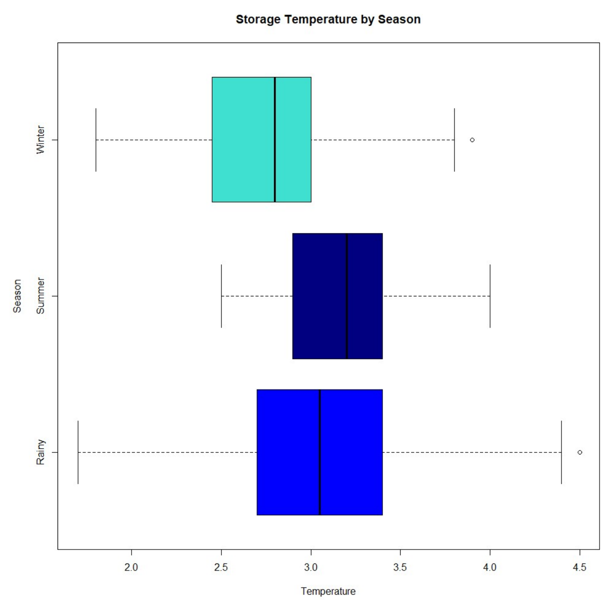
One way ANOVA properties tests:
1- Test of Normality: Since p-value of the test is larger than 5% significance level, we fail to reject the null hypothesis, that the data follows the normal distribution.

Problem 1 – Question7

Are all season temperature mean equal :

0.05044
P-Value Normality

0.0003951
P-Value Homogeneity of variance

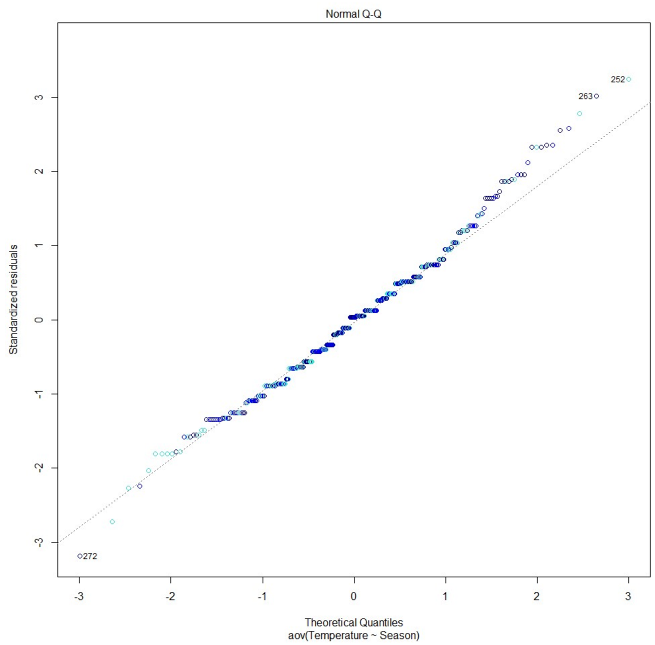


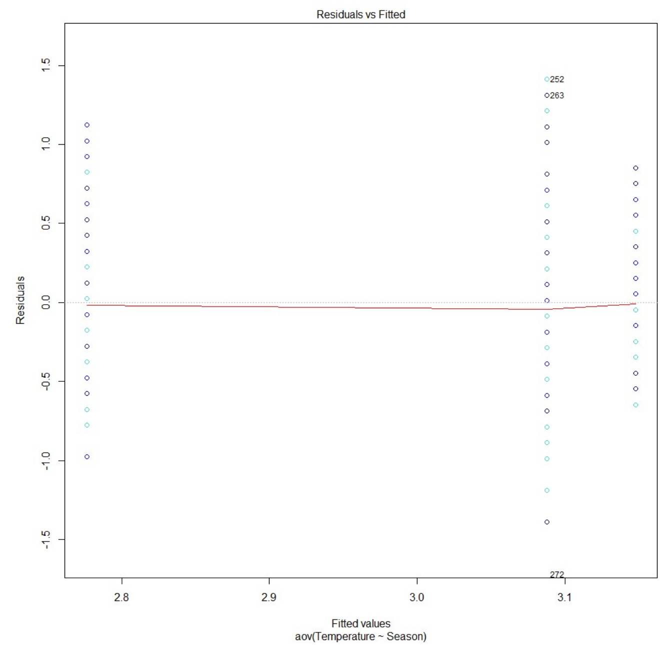
One way ANOVA properties tests:
2- Test of Homogeneity of variance :Since the p-value is smaller than 5% significance level, we reject the null hypothesis of homogeneity of variances.

Problem 1 – Question7

Are all season temperature mean equal :

2 **Df**	25.32 **F-Stat**	5.08e-11 **P-Value**



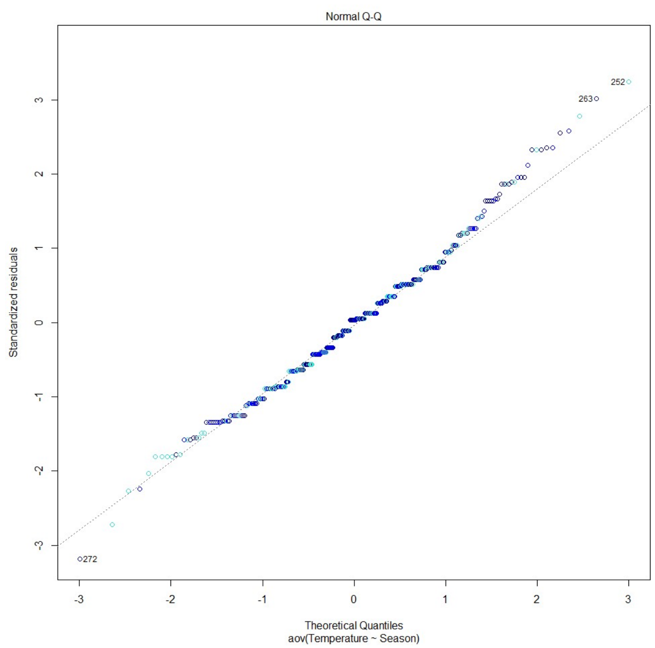


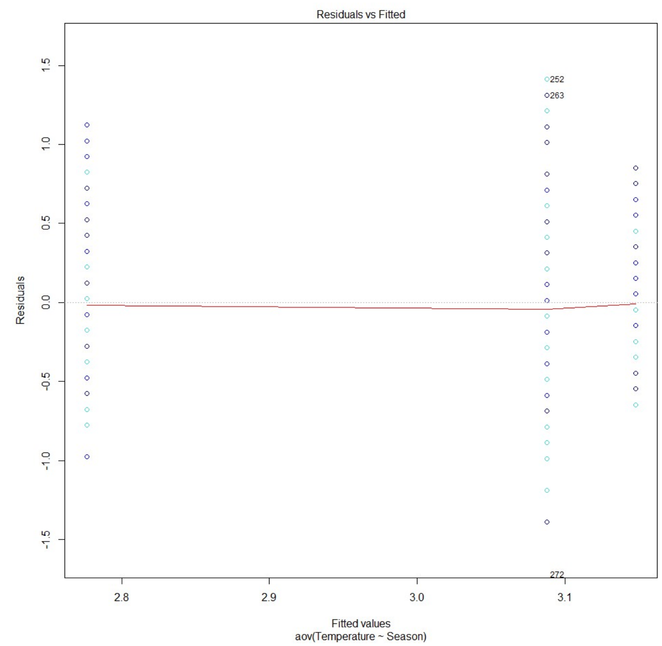
The P-value of the ANOVA test is 5.08e-11, which is highly significant, therefore we reject the null hypothesis that the three population means are identical. At least for one season mean temperature is different from the rest.

Problem 1 – Question7

Are all season temperature mean equal :

2 **Df**	25.32 **F-Stat**	5.08e-11 **P-Value**



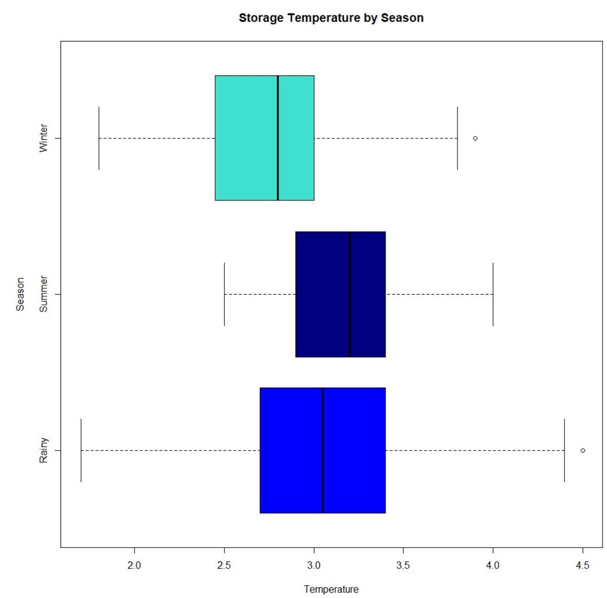


Residuals vs Fitted graph show that two seasons means are close to each other, while the third is faraway.
Normal Q-Q graph show indicates that the normality assumption holds.

Problem 1 – Question7

Are all season temperature mean equal :

2.5368e-12
P-Value
Welch t.test



So, in case there is no homogeneity of variances, we use Welch t.test to confirm ANOVA Results.
The welch test results support ANOVA results, we can confirm that at least one of the three populations means is different

Problem 1 – Question7

Are all season temperature mean equal :

Summer-Rainy	Winter-Rainy	Winter-Summer
0.5376924	0.0000002	0.0000000
P-Value	**P-Value**	**P-Value**
Paired t.test	**Paired t.test**	**Paired t.test**

P-value is significant for comparing temperature mean levels for these paired seasons winter rainy and winter-summer, but not for Summer-rainy. The null hypothesis of equality of all population means is rejected. It is now clear that mean temperature for Summer and rainy is similar but temperature for type winter is significantly different from these two.

Problem 2.

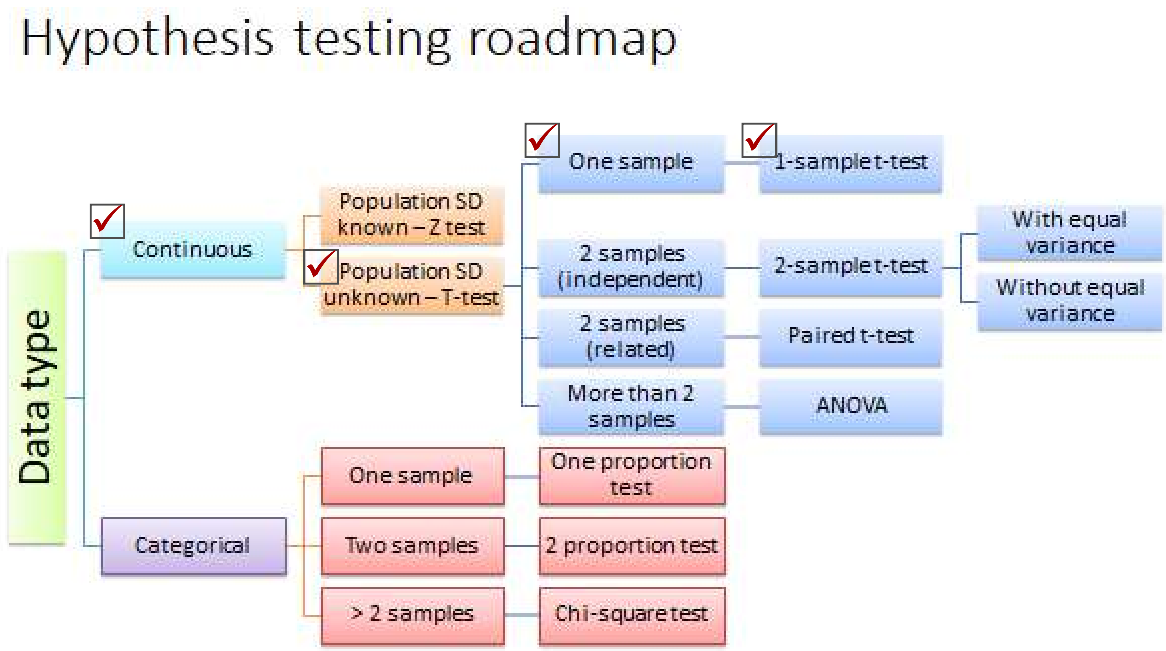
In Mar 2018, Cold Storage started getting complaints from their clients that they have been getting complaints from end consumers of the dairy products going sour and often smelling. On getting these complaints, the supervisor pulls out data of last 35 days' temperatures. As a safety measure, the Supervisor decides to be vigilant to maintain the temperature at 3.9 C or below.

Assume 3.9 C as the upper acceptable value for mean temperature and at alpha = 0.1. Do you feel that there is a need for some corrective action in the Cold Storage Plant or is it that the problem is from the 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. (8 marks)*
2. *State the Hypothesis, perform hypothesis test and determine p-value (11 marks)*
3. *Give your inference (4 marks)*

Problem 2 – Question1

Type of Hypotheses test :

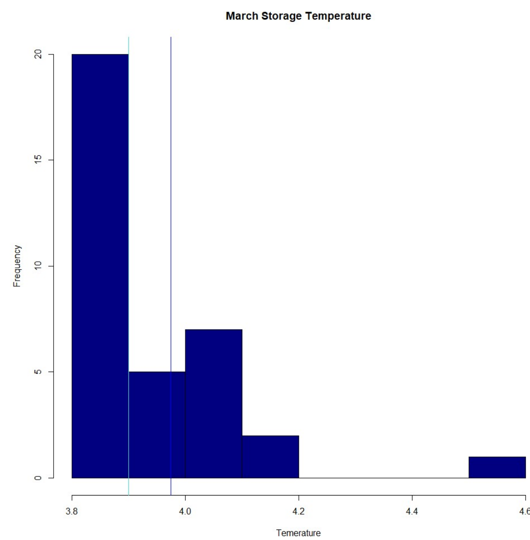


As we trying to test the mean of a single group against a known mean, hence we should use one sample t-test, single tail. We chose the t-test as the population parameters (mean and standard deviation) are not known.

Problem 2 – Question2

Hypotheses test for cold storage corrective action:

H0: Storage Temp <= 3.9 C
Ha: Storage Temp > 3.9 C



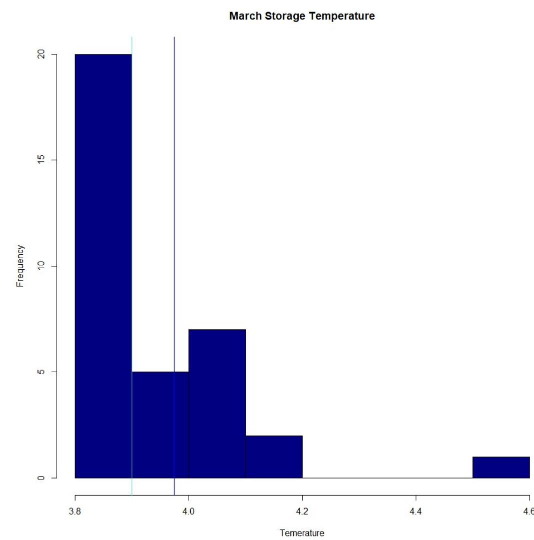
March Temperature data seems to be skewed to the right, meaning on considerably less days the Temperature surpassed a mean of 3.974 C.

Problem 2 – Question2

Hypotheses test for cold storage corrective action:

3.974C
March Average Temp

0.15967C
March SD Temp

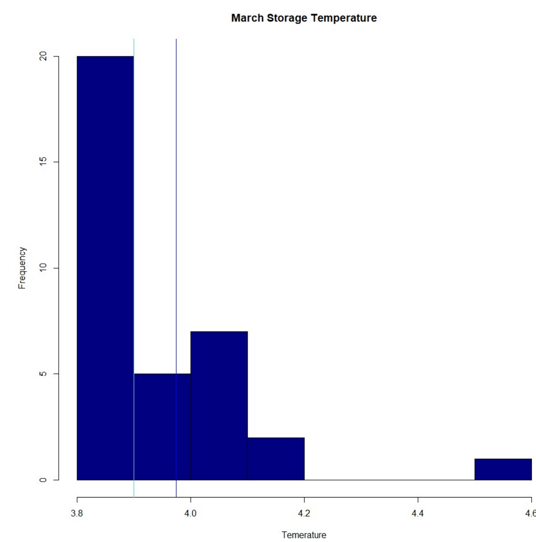


March Temperature data Temperature data seems vary from the mean by ~ 0.16C from the mean temp ~ 3.974C.

Problem 2 – Question2

Hypotheses test for cold storage corrective action:

34 **Df**	2.75 **T-Stat**	0.004711 **P-Value**

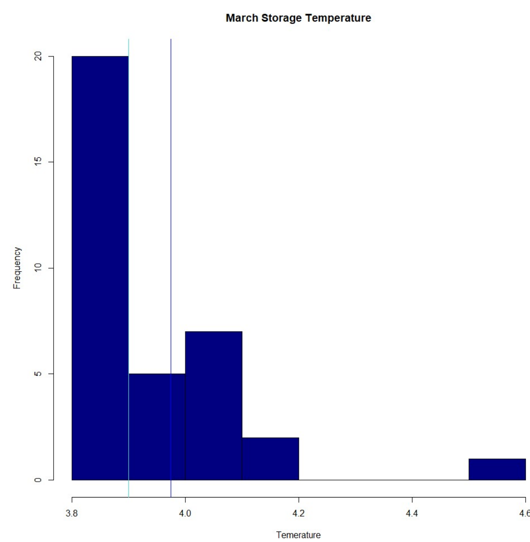


The T-test P-value amounts to 0.004711.

Problem 2 – Question3

Hypotheses test Inference:

34 **Df**	2.75 **T-Stat**	0.004711 **P-Value**



Inference:
We reject the null hypothesis, statistical evidence shows average temperature is very likely to exceed 3.9 C. So, a corrective action in the Cold Storage Plant is required.

R Code

Project 2

Mohanned Gomaa

11/8/2019

Module 2 - Project 2 Cold Storage Problem

Set a Working directory, for ease of processing.

```
setwd('C:/Users/mgomaa032/Desktop/PwC/Projects/DXB/PnO/WorkFroce
analytics/UAT/Module 2/Project') # Check that WD was set. getwd()
```

```
#Import data from CSV file TemperatureData =
read.csv('Cold_Storage_Temp_Data.csv',header = T)
```

View Data

```
View(TemperatureData) head(TemperatureData) tail(TemperatureData)
```

attach data for easy access

```
attach(TemperatureData)
```

Quick overview of data

```
summary(TemperatureData) str(TemperatureData)
```

Visualize_Data

```
hist(Temperature,main = 'Storage Temperature', col = 'Navy') abline(v=
mean(Temperature), col='Blue') abline(v= median(Temperature), col='turquoise')
#Temperature data seems to normally distributed, no visible signs of skewness
boxplot(Temperature~Season, horizontal = T, main='Storage Temperature by Season',
col=c('Blue', 'Navy','turquoise')) #Temperature data by season seems to be of a similar
distribution, with small expectation of winter.
```

R Code

#Problem1 #Q1: Mean Temperature by Season

Avg_Temp_Session=by(Temperature,INDICES = Season, FUN = mean) Avg_Temp_Session # Visualize Data Summer= hist(TemperatureData*Temperature* [*TemperatureData* Season==“Summer”],main = ‘Storage Temperature - Summer’, col = ‘Navy’,xlab = ‘Temp’) abline(v= mean(TemperatureData*Temperature* [*TemperatureData* Season==“Summer”]), col=‘Blue’) abline(v= median(TemperatureData*Temperature* [*TemperatureData* Season==“Summer”]), col=‘turquoise’)

Winter= hist(TemperatureData*Temperature* [*TemperatureData* Season==“Winter”],main = ‘Storage Temperature - Winter’, col = ‘Navy’,xlab = ‘Temp’) abline(v= mean(TemperatureData*Temperature* [*TemperatureData* Season==“Winter”]), col=‘Blue’) abline(v= median(TemperatureData*Temperature* [*TemperatureData* Season==“Winter”]), col=‘turquoise’)

Rainy= hist(TemperatureData*Temperature* [*TemperatureData* Season==“Rainy”],main = ‘Storage Temperature - Rainy’, col = ‘Navy’,xlab = ‘Temp’) abline(v= mean(TemperatureData*Temperature* [*TemperatureData* Season==“Rainy”]), col=‘Blue’) abline(v= median(TemperatureData*Temperature* [*TemperatureData* Season==“Rainy”]), col=‘turquoise’)

#Q2: Overall mean for the full year # Assumption: As no year column was inculded in the data set, # I have assumed that data belong to one year. Avg_Temp_Annual=mean(Temperature) Avg_Temp_Annual

#Q3: Overall Standard Deviation for the full year # Assumption: As no year column was inculded in the data set, # I have assumed that data belong to one year. SD_Temp_Annual = sd(Temperature) SD_Temp_Annual

Q4: The probability of temperature have fallen below 2 C

Assumption: Normal distribution assumed

Temp_Prop_lessthan_2 = pnorm (2, mean = Avg_Temp_Annual, sd = SD_Temp_Annual, lower.tail = T)

Q5: The probability of temperature have gone above 4 C

Assumption: Normal distribution assumed

Temp_Prop_Morethan_4 = pnorm (4, mean = Avg_Temp_Annual, sd = SD_Temp_Annual, lower.tail = F) Temp_Prop_Morethan_4 0.01612075*100

R Code

Q6: AMC Company penalty

P(2.5%<Propability of Temp<5%)=10%

Temp_Prop_lessthan_2+Temp_Prop_Morethan_4

#Q7: One way Annova # Assumption:Seasons are the factors # H0: Temp_Winter=Temp_Rainy= Temp_Summar # Ha: H0 is not true # Test for normailty Is_it_normal=shapiro.test(Temperature) Is_it_normal 0.05044>0.05 # Since p-value of the test is larger than 5% significance level, # we fail to reject the null hypothesis that the response follows the normal distribution.

Test for Homogeneity of varaince require(Rcmdr) #call library Are_Var_equal=leveneTest(Temperature~Season,data = TemperatureData) Are_Var_equal 0.0003951>0.05 # Since the p-value is smaller than 5% significance level, # we reject the null hypothesis of homogeneity of variances.

Average Temperature by session Mean_Temp_By_Session= by(Temperature,INDICES = Season,FUN = mean) Mean_Temp_By_Session # At first glance, mean seem to with similar range, with minor expectation of winter.

Varaince Temperature by session Var_Temp_By_Session= by(Temperature,INDICES = Season,FUN = var) Var_Temp_By_Session #Variance seems within an approximate range, with exception to rainy session.

#Anova Model Anova_Temp= aov(Temperature~Season,data = TemperatureData) summary(Anova_Temp) # The Pvalue is 5.08e-11, which is highly significant, therefore we reject # the null hypothesis that the three population means are identical. # At least for one season mean temperature is different from the rest.

#Anova ModelSupport Graphs # The following two graphs are introduced to check the distribution of the residuals. plot(Anova_Temp, col=c('Blue', 'Navy','turquoise')) # Residuals vs Fitted graph show that two seasons means are close to each other,while the third is faraway. # Normal Q-Q graph show indicates that the normality assumption holds.

Alternative for Anova, when varaince is unequal

install.packages('onewaytests') require(onewaytests) Unequal_var_test= welch.test(Temperature~Season,data = TemperatureData) Unequal_var_test ?welch.test # The welch test results support ANOVA results, we can confirm # that at least one of the three populations means is different

#Paired Test TukeyHSD(Anova_Temp) # P-value is significant for comparing temperature mean levels for these paired seasons winter rainy and winter-summer, # but not for Summer-rainy. The null hypothesis of equality of all # population means is rejected. It is

R Code

now clear that mean temperature # for Summer and rainy is similar but temperature for # type winter is significantly different from these two.

#Problem2 #Ho: Storage Temp <= 3.9 C #Ha: Storage Temp > 3.9 C # a=0.01

#Import data from CSV file MarTemperatureData = read.csv('Cold_Storage_Mar2018.csv',header = T)

#View Data View(MarTemperatureData) head(MarTemperatureData) tail(MarTemperatureData)

#attach data for easy access attach(MarTemperatureData)

#Quick overview of data summary(MarTemperatureData) str(MarTemperatureData)

#Visualize_Data hist(MarTemperatureData$Temperature, main = 'MarchStorageTemperature', col = 'Navy', xlab = 'Temerature')abline(v = mean(MarTemperatureData$ Temperature), col='Blue') abline(v= median(MarTemperatureData$Temperature), col='turquoise') # March Temperature data seems to be skewed to the right, # meaning on considerably less days the Temperature surpassed a mean of 3.974 C.

Q1: Type of hypothesis test

As we trying to test the mean of a single group against

a known mean, hence we should use one sample t-test, single tail.

We chose the t-test as the population parameters (mean and standard deviation)

are not known.

#Q2: Hypothesis test

#Setp 1: State Ho and Ha # Ho: Storage Temp <= 3.9 C # Ha: Storage Temp > 3.9 C # a=0.01

Step1: Avg. Storage Temperature or mu

March_Avg_Temp= mean(MarTemperatureData$Temperature) March_Avg_Temp

R Code

Step2: Sd Storage Temperature

March_sd_Temp= sd(MarTemperatureData$Temperature) March_sd_Temp

Step3: Calculate P-value

Hypo_test = t.test(MarTemperatureData$Temperature,mu=3.9,alternative = ‘greater’,conf.level = 0.99) Hypo_test t.stat = (March_Avg_Temp - 3.9)/(March_sd_Temp/sqrt(35)) t.stat Pvalue= 1- pt(t.stat, 35-1) Pvalue

Q3: Inference

0.004711<0.01 # We reject the null hypothesis, statistical evidence shows average temperature # is very likely to exceed 3.9 C.So, a corrective action in the Cold Storage Plant is required.

Thank You