Assessment: Project 2

Cold Storage Problem

Thangamani R

Contents

Problem 1:

1. Mean cold storage temperature for each season:

Find mean cold storage temperature for Summer, Winter and Rainy Season (3 marks)

a. Set working directory
b. Load the csv file
c. View the summary of the csv file.

```
setwd("C:\\Users\\thang\\OneDrive\\DSBA\\Week7")
cold_storage_temp=read.csv("Cold_Storage_Temp_Data.csv");
summary(cold_storage_temp)
```

```
> cold_storage_temp=read.csv("Cold_Storage_Temp_Data.csv");
> summary(cold_storage_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
>
```

1. Rainy Season

Mean cold storage temperature for rainy season is 3.087705

```
#1.Mean Temp for rainy season
print(mean(subset(cold_storage_temp,Season=='Rainy')[,'Temperature']))
```

```
> #1.Mean Temp for rainy season
> print(mean(subset(cold_storage_temp,Season=='Rainy')[,'Temperature']))
[1] 3.087705
>
```

2. Winter Season

Mean cold storage temperature for winter season is 2.776423

```
#2.Mean Temp for Winter season
print(mean(subset(cold_storage_temp,Season=='Winter')[,'Temperature']))
```

```
> #2.Mean Temp for Winter season
> print(mean(subset(cold_storage_temp,Season=='Winter')[,'Temperature']))
[1] 2.776423
>
```

3. Summer Season

Mean cold storage temperature for summer season is 3.1475

```
#3.Mean Temp for Summer season
print(mean(subset(cold_storage_temp,Season=='Summer')[,'Temperature']))
```

```
> print(mean(subset(cold_storage_temp,Season=='Summer')[,'Temperature']))
[1] 3.1475
>
```

2. Mean cold storage temperature for the entire year

Find overall mean for the full year (3 marks)

Overall mean for the full year is 3.002466

```
#4.Mean Temp for entire year
print(mean(cold_storage_temp[,'Temperature']))
```

```
> #4.Mean Temp for entire year
> print(mean(cold_storage_temp[,'Temperature']))
[1] 3.002466
>
```

3. Standard Deviation for the full year

Standard deviation for the full year is 0.4658319

```
#5.Standard deviation of Temp for whole year
print(sd(cold_storage_temp[,'Temperature']))
```

```
> print(sd(cold_storage_temp[,'Temperature']))
[1] 0.4658319
>
```

4. Probability of temperature falling below 2 C

As it is below 2 C use the *lower.tail=T*

Use the mean, standard deviation of the whole year to compute pnorm. The computed value of pnorm is 0.01569904

```
#6.Probability of temp being below 2 Degrees
pnorm(2,3.002466,0.4658319,lower.tail=T)
```

```
> pnorm(2,3.002466,0.4658319,lower.tail=T)
[1] 0.01569904
>
```

5. Probability of temperature having gone above 4 C

There are two ways by which this can be done either using pnorm with lower.tail=F or using 1-pnorm with lower.tail=T.

Both the methods are shown in the screen shot below

```
#7. Probability of temp being greater than 4 degress
1-pnorm(4,3.002466,0.4658319,lower.tail=T)

pnorm(4,3.002466,0.4658319,lower.tail=F)
```

And the calculated value is 0.01612076

```
> #7. Probability of temp being greater than 4 degress
> 1-pnorm(4,3.002466,0.4658319,lower.tail=T)
[1] 0.01612076
> pnorm(4,3.002466,0.4658319,lower.tail=F)
[1] 0.01612076
>
```

6. Penalty for the AMC company:

Probability of the temp being less than 2 degrees = 0.01569904

Probability of the temp being greater than 4 degrees = 0.01612076

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

But P$(A \cap B)$ is zero as the probabilities are mutually exclusive.

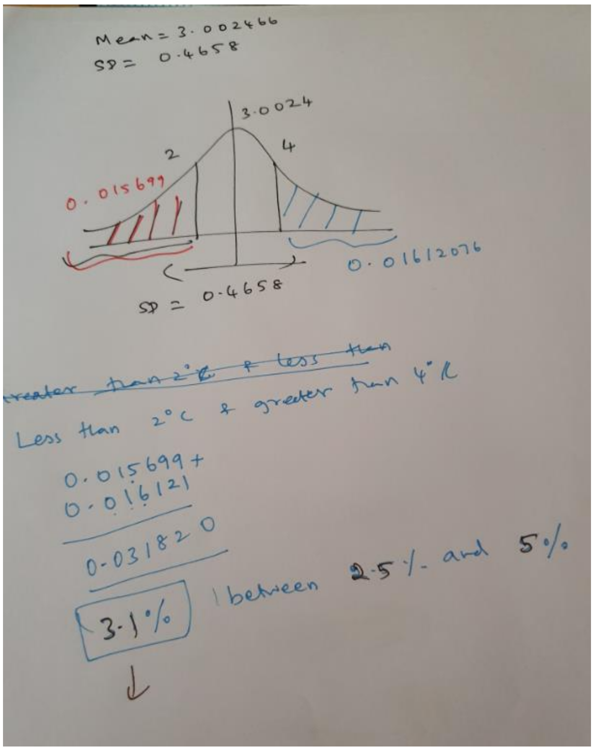
i.e less than 2 C and greater than 4 C cannot occur simultaneously.

Hence sum of the probabilities

$P(A) + P(B)$= 0.0318198

Which is 3.1%.

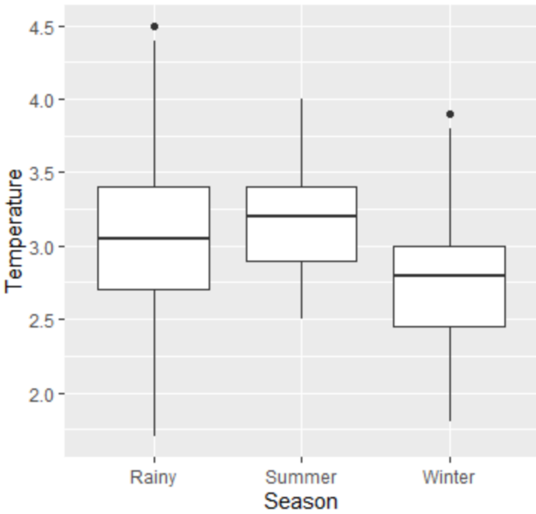
As 3.1% falls between 2.5% and 5% the penalty for the AMC company would be <u>10% of the AMC(Annual Maintenance Contract)</u>



7. Perform one-way ANOVA Test:

- Plotting the temperatures of the various seasons using boxplot

```
ggplot(cold_storage_temp,aes(x=Season,y=Temperature))+
  geom_boxplot()+scale_x_discrete()
```



- From the box plot it appears that the mean during the 3 seasons are well within the 2-4 C range, but the means are not equal.
- Upon executing the one-way ANOVA calculations against Temperature and Season columns in cold storage data set the result is as follows

```
sum_data=aov(Temperature~Season,data=cold_storage_temp)
summary(sum_data)
```

```
> sum_data=aov(Temperature~Season,data=cold_storage_temp)
> summary(sum_data)
             Df Sum Sq Mean Sq F value   Pr(>F)
Season        2   9.70   4.848   25.32 5.08e-11 ***
Residuals   362  69.29   0.191
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

- P value is way below 0.05 hence the null hypothesis is rejected.
- Status quo is temperatures are maintained across all the seasons between 2- 4 C which is the null hypothesis.
- Alternate hypothesis is that the temperatures are not maintained across seasons between 2 C-4 C. Temperatures increase and hence the dairy products are spoiled.
- As the null hypothesis is rejected, alternate hypothesis is true.
- Hence the conclusion is that temperatures are not maintained uniform across all seasons.
- To further find the difference in mean across seasons we run the Tukey test

```
#P is lower than 0.05 then reject null hypothesis
#status quo is temp are maintained, but this is rejected and the conclusion
#is that temperaturs vary and are not between 2-4 always
tukey_result=TukeyHSD(sum_data)
tukey_result
```

```
> tukey_result=TukeyHSD(sum_data)
> tukey_result
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Temperature ~ Season, data = cold_storage_temp)

$Season
                    diff         lwr        upr     p adj
Summer-Rainy   0.05979508 -0.07258434  0.1921745 0.5376924
Winter-Rainy  -0.31128215 -0.44284519 -0.1797191 0.0000002
Winter-Summer -0.37107724 -0.50318954 -0.2389649 0.0000000

>
```

- From Tukey tests we infer that:
    - Between Summer & Rainy seasons, the p-vale is 0.5 which is greater than 0.05 hence null hypothesis cannot be rejected. i.e. no significant change in temperatures in Summer and Rainy seasons.
    - Between winter and Rainy seasons, the p-value is 0 which is less than 0.05 and hence null hypothesis can be rejected as the p value is low. So, there is significant difference in the mean of the temperatures between Winter and Rainy seasons.

   - Between winter and summer seasons, the p-value is 0 which is less than 0.05 and hence null hypothesis can be rejected as the p value is low. So, there is significant difference in the mean of the temperatures between Winter and summer seasons.
- As there is at least one difference in the means we reject the null hypothesis.
- So, the alternate hypothesis is true, i.e. the cold storage temperature is not maintained uniform between 2 C -4 C across all seasons.

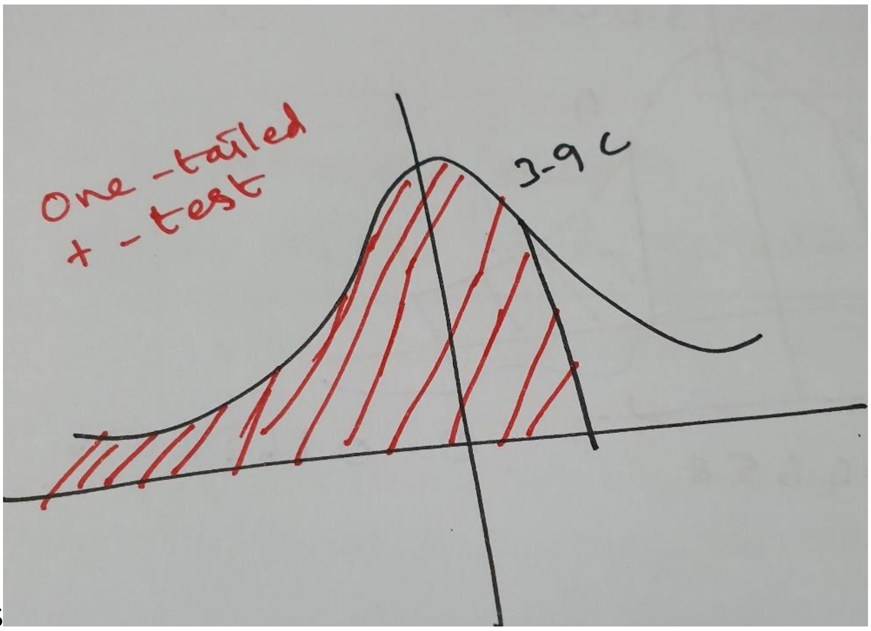
Problem 2:

1. Hypothesis Test:

We need to measure that the maximum acceptable upper temperature is 3.9 C

- Standard deviation is not known, so z test cannot be done. It should be a T test.
- Only one sample is being measured. i.e. the temperature in Summer only and hence it is one sample T test.
- It is **not ANOVA** as we are not comparing temperatures across various seasons, i.e. multiple samples like during summer, winter, rainy season are not involved.
- Only the lower range up to 3.9 C need to be measured
- We need to do **one tailed T test** as only the lower range needs to be measured**.**
- Given alpha =0.1, so the confidence level =0.9
- Null Hypothesis: Temperature is less than or equal to 3.9 C
- Alternate Hypothesis: Temperature exceeds 3.9 C

S



2. Hypothesis :

Null Hypothesis: Null Hypothesis is that the temperature of the cold storage is less than or equal to 3.9 C always.

Alternate Hypothesis: Temperature of the cold storage is greater than 3.9 C

Hypothesis Test:

- Load the March dataset
- Summary of the March dataset.
- Summary shows that the mean temperature is greater than 3.9 C

```
#Part 2

#One tailed
# No standard deviation is known, so t-test
#alpha=0.1=> Confidence level=0.9
#mean =3.9
cold_storage_march=read.csv("Cold_Storage_Mar2018.csv");
summary(cold_storage_march)
```

```
> summary(cold_storage_march)
    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
>
```

- Do the one tailed t-test as discussed in the previous section

```
t.test(cold_storage_march$Temperature,conf.level = 0.9,alternative = 'greater'
       ,paired=FALSE)
```

```
> t.test(cold_storage_march$Temperature,conf.level = 0.9,alternative = 'greater'
+        ,paired=FALSE)

        One Sample t-test

data:  cold_storage_march$Temperature
t = 147.25, df = 34, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0
90 percent confidence interval:
 3.939011      Inf
sample estimates:
mean of x
 3.974286
```

- P-value is less hence null should go. Hence, we reject the null hypothesis.
- i.e. the null hypothesis is not true
- Conclusion is that the temperature of cold storage is not always less than or equal to 3.9 C.
- It goes beyond 3.9 C also
- Hence the alternate hypothesis is true.

3. Inference:

As seen from above two steps in problem 2 , the temperature of the cold storage is not always less than or equal to 3.9 C. Sometimes the temperature goes beyond 3.9 C as well. Due to this there are chances of the dairy products getting spoiled. Hence corrective action is required from the cold storage to monitor the temperature regularly and enough stringent measures taken to make sure that the temperature is maintained less than or equal to 3.9 C. This will avoid the dairy products being spoiled.