Mini Project 1 - Cold Storage Case Study

Submitted by

Sita K

BACP Batch (Dec 19-May 20)

Great Learning

January 19, 2020

Table of Contents

1. Project Objective

The objective of this report is to explore the cold data storage datasets (Cold_Storage_Mar2018 and Cold_Storage_Temp_Data) in R and generate insights about the data set. This report will contain:

- Mean and standard deviation of the data set
- Assuming normal distribution, application of descriptive statistics
- Calculation of probability of certain conditions
- Applying hypothesis test for an alternative hypothesis
- Final conclusion with the output of the test

2. Assumptions

- The data provided is conclusive and contains the required data

3. Data Analysis – Approach

1. Environment data setup and data import
2. Calculating the required values using inbuilt functions
3. Determining the hypothesis to be used
4. Conduct the hypothesis test
5. Form the conclusion

For environment data setup, R's inbuilt packages were used. Also for setting up working directory '**setwd**()' function was used. The given dataset is in .csv format, so we can use **read.csv** function to import the data. All the R commands are in Appendix A.

4. Problem 1 Responses

1. The data set contains temperatures in Summer, Winter and Rainy Seasons. So to find the mean temperature in each season, we have to group the temperatures by season and find the mean of each group. I used the **aggregate** function in R, to group the temperature values and find individual mean. The answers are:
   Rainy – 3.039
   Summer – 3.153
   Winter - 2.701
2. For finding the overall mean of the temperature for the full year, we can use the **mean** function in R.
   The answer is 2.963
3. To find the standard deviation of the temperature for the full year, we can use **sd** function in R.
   The answer is 0.509
4. Assuming normal distribution, to find the probability of temperature having fallen below 2 C, we can use the **pnorm** function in R
   The answer is .029 i.e. 2.9%

5. Assuming normal distribution, to find the probability of temperature having gone above 4 C, we can use the **pnorm** function in R
   The answer is 0.021 i.e. 2.1%
6. The penalty is 10% of AMC, if the probability of temperature falling below 2C and above 4 C is between 2.5% and 5%. And the penalty is 25% AMC if the probability exceeded 5%.
   Probability of temperature falling below 2C = 2.9%
   Probability of temperature going above 4C = 2.1%
   Since one of the probabilities are falling between 2.5% and 5%, the penalty is 10% of AMC.

5. Problem 2 Responses

1. Here H0 - $\mu = 3.9C$ – Zero hypothesis
   Ha - $\mu > 3.9C$ – Alternative hypothesis
   $\alpha=0.1$
   Since it's a small sample data and standard deviation is not given for the entire population, we have to use one-tailed/right tailed T test since the mean value can be only on the higher side according to alternative hypothesis to confirm corrective action is needed in the Cold storage plant.
2. Hypothesis is H0 - $\mu = 3.9C$ – Zero hypothesis
   Ha - $\mu > 3.9C$ – Alternative hypothesis
   $\alpha=0.1$
   We can do one tailed T test in R using the function **t.test**. Since the confidence level is specifically mentioned as 0.9 not 0.95 which is the default we have to mention the same in t.test function and also alt has to be given as 'greater' since this is a right tailed T test.
   The answer of p is 0.047
3. Since the p value is 0.047 which is less than the alpha value i.e. 0.1 we can reject the zero hypothesis and thus alternative hypothesis is true.
   And mean temperature is greater than 3.9C and corrective action is needed in the storage plant.

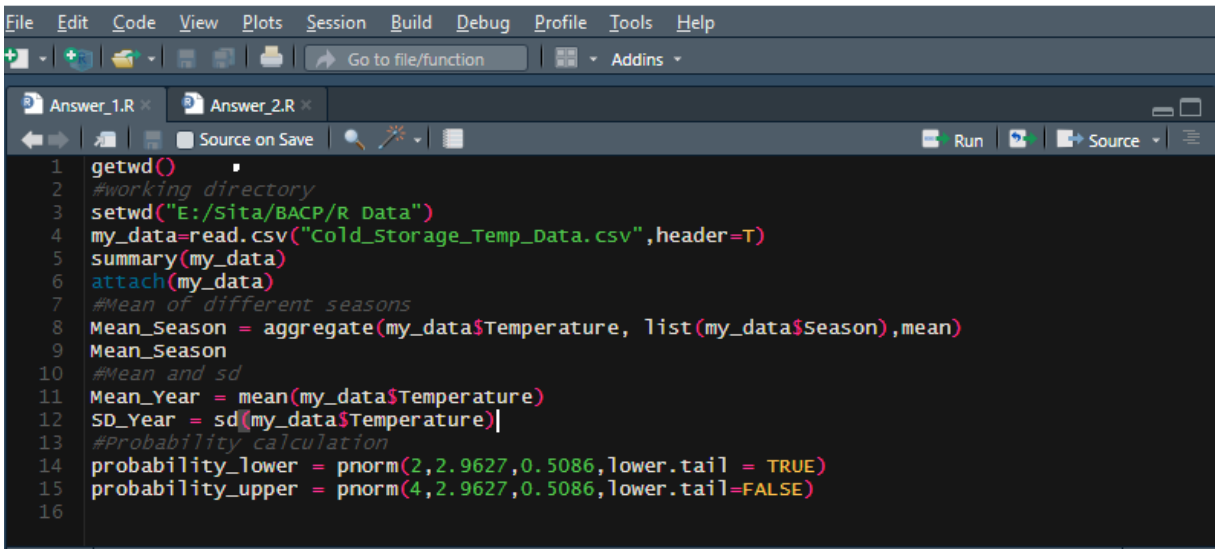
6. Conclusion

Mean temperature is greater than 3.9C and corrective action is needed in the storage plant.

Appendix A

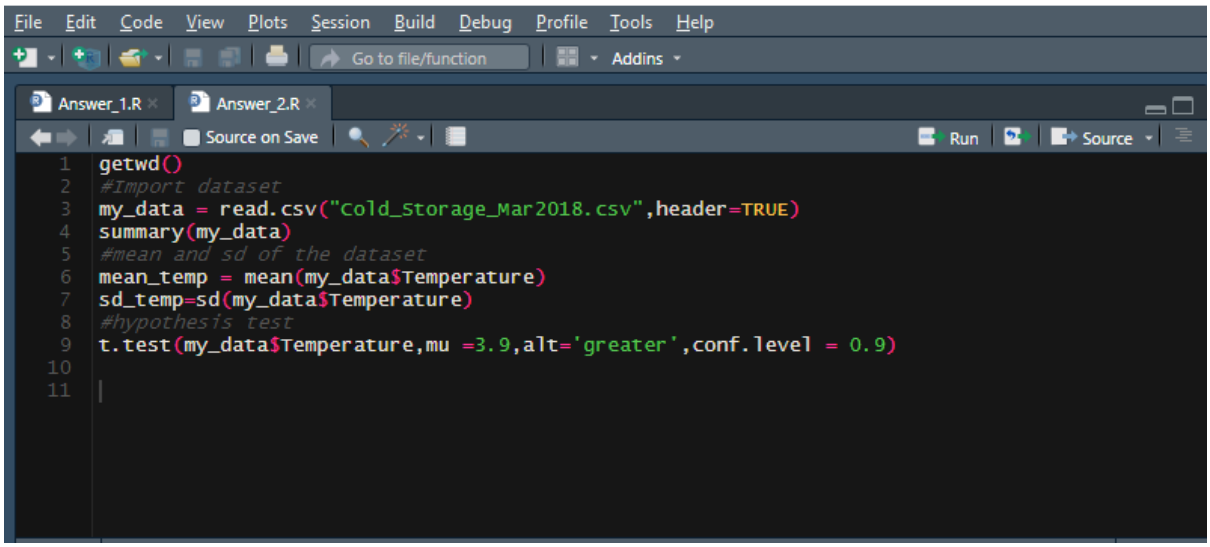
Answer_1.R and Answer_2.R contains the responses. Sample code is given below.

RStudio



```
getwd()
#working directory
setwd("E:/Sita/BACP/R Data")
my_data=read.csv("Cold_Storage_Temp_Data.csv",header=T)
summary(my_data)
attach(my_data)
#Mean of different seasons
Mean_Season = aggregate(my_data$Temperature, list(my_data$Season),mean)
Mean_Season
#Mean and sd
Mean_Year = mean(my_data$Temperature)
SD_Year = sd(my_data$Temperature)
#Probability calculation
probability_lower = pnorm(2,2.9627,0.5086,lower.tail = TRUE)
probability_upper = pnorm(4,2.9627,0.5086,lower.tail=FALSE)
```

RStudio



```
getwd()
#Import dataset
my_data = read.csv("Cold_Storage_Mar2018.csv",header=TRUE)
summary(my_data)
#mean and sd of the dataset
mean_temp = mean(my_data$Temperature)
sd_temp=sd(my_data$Temperature)
#hypothesis test
t.test(my_data$Temperature,mu =3.9,alt='greater',conf.level = 0.9)
```