STATISTICAL METHODS FOR DECISION MAKING

GROUP ASSIGNMENT

(GROUP-9)

Done by:

1. Akshay Subramanian
2. Gokul Chinnasamy
3. Jashwanth Govardhan
4. Mohamed Yusuf
5. Nithya

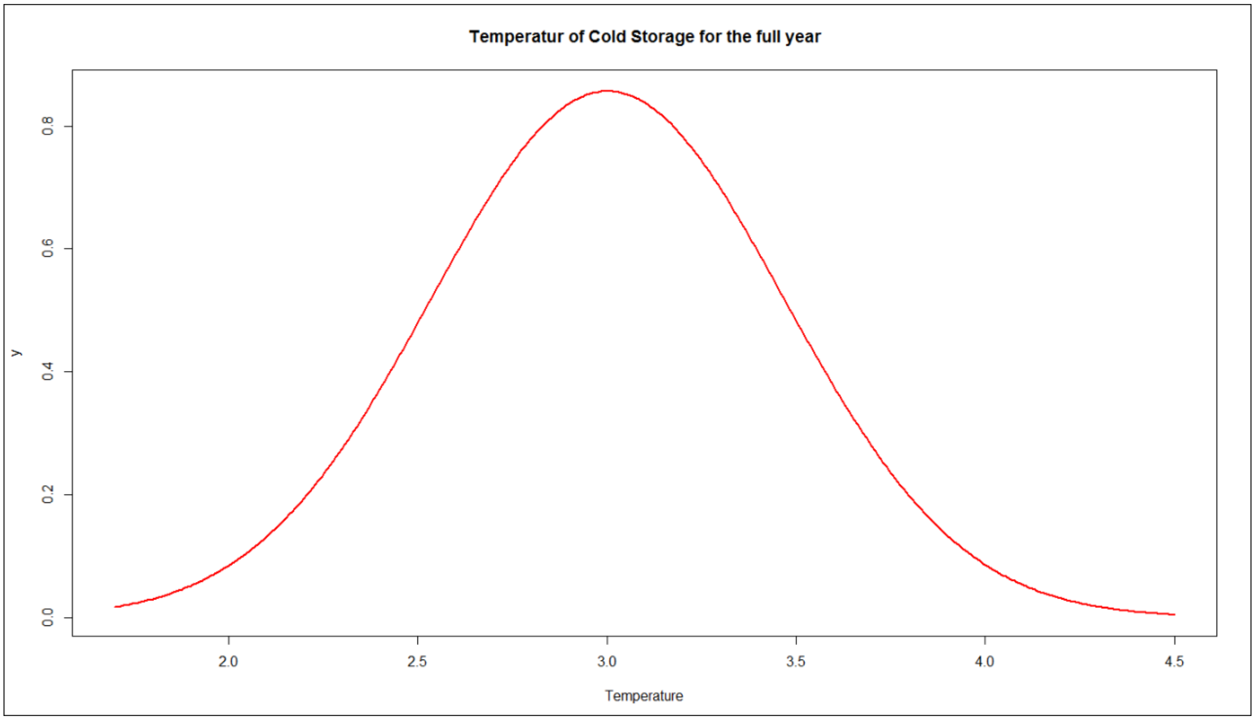
Submitted on: 20/10/2019.

Tool used: R Studio

Problem 1:

On analyzing the given datasheet, it is found that the data has been recorded for 12 months and 365 days in a year across summer, winter and rainy seasons. It is understood that, the temperature range should not exceed more than or less than the fixed optimal temperature range of 2^o C – 4^o C which would not spoil the quality of the Cold Storage products Pasteurized Fresh Whole or Skimmed Milk, Sweet Cream, Flavored Milk Drinks.

The graph obtained by plotting the Cold Storage temperatures which is Normally Distributed for the full year in R Studio:



1. Find the mean cold storage temperature for summer, winter and rainy seasons.

   Solution:

   For obtaining the mean values for every season, the season should be mentioned as a subset/filtered option in the mean command used in R Studio.

   For finding the mean of temperatures for "Summer" season:

```
summer <- mean(df_coldstg$Temperature[df_coldstg$Season=="Summer"])
print(summer)
```

   The value obtained is:

```
> print(summer)
[1] 3.1475
```

For finding the mean of temperatures for “Winter” season:

```
winter <- mean(df_coldstg$Temperature[df_coldstg$Season=="Winter"])
print(winter)
```

The value obtained is:

```
> print(winter)
[1] 2.776423
```

For finding the mean of temperatures for “Rainy” season:

```
rainy <- mean(df_coldstg$Temperature[df_coldstg$Season=="Rainy"])
print(rainy)
```

The value obtained is:

```
> print(rainy)
[1] 3.087705
```

2. Find overall mean temperature for the full year.

<u>Solution:</u>

For finding the mean temperature for the whole year, the “mean” command is used without any filtering conditions.

```
mean(df_coldstg$Temperature)
```

The value obtained is:

```
> mean(df_coldstg$Temperature)
[1] 3.002466
```

3. Find Standard deviation for the full year.

<u>Solution:</u>

For finding the standard deviation for the whole year, the command “sd” is used.

```
sd(df_coldstg$Temperature)
```

The value obtained is:

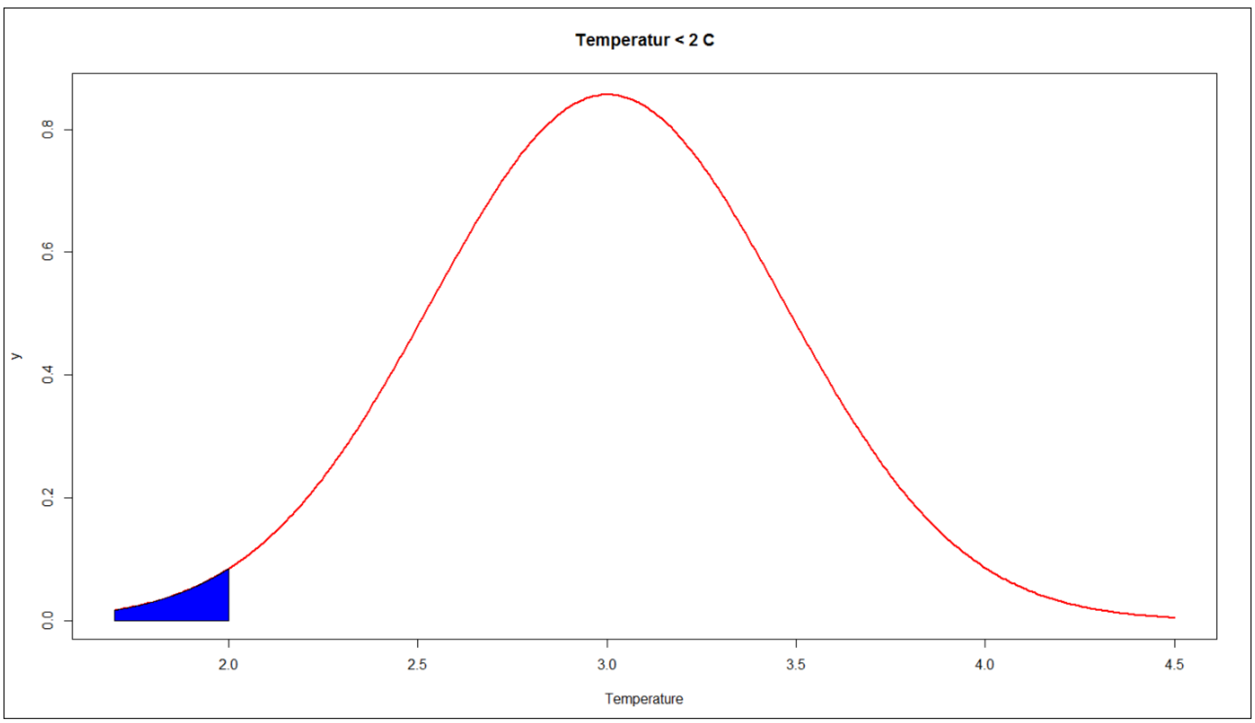
```
> sd(df_coldstg$Temperature)
[1] 0.4658319
```

On looking into the graph plotted with the given values, we assume that the temperature values are normally distributed.

4. Assuming normal distribution, find the probability of the temperature fallen below 2ºC.

<u>Solution:</u>

To find: Probability (Temperature<2ºC)



Steps:

- Find the Z value for 2ºC using the formula Z=(X-μ)/σ
- To find the probability for the Z value, the command "pnorm" is used.

```
X <- (2-3.002466)/0.4658319
pnorm(X)
```

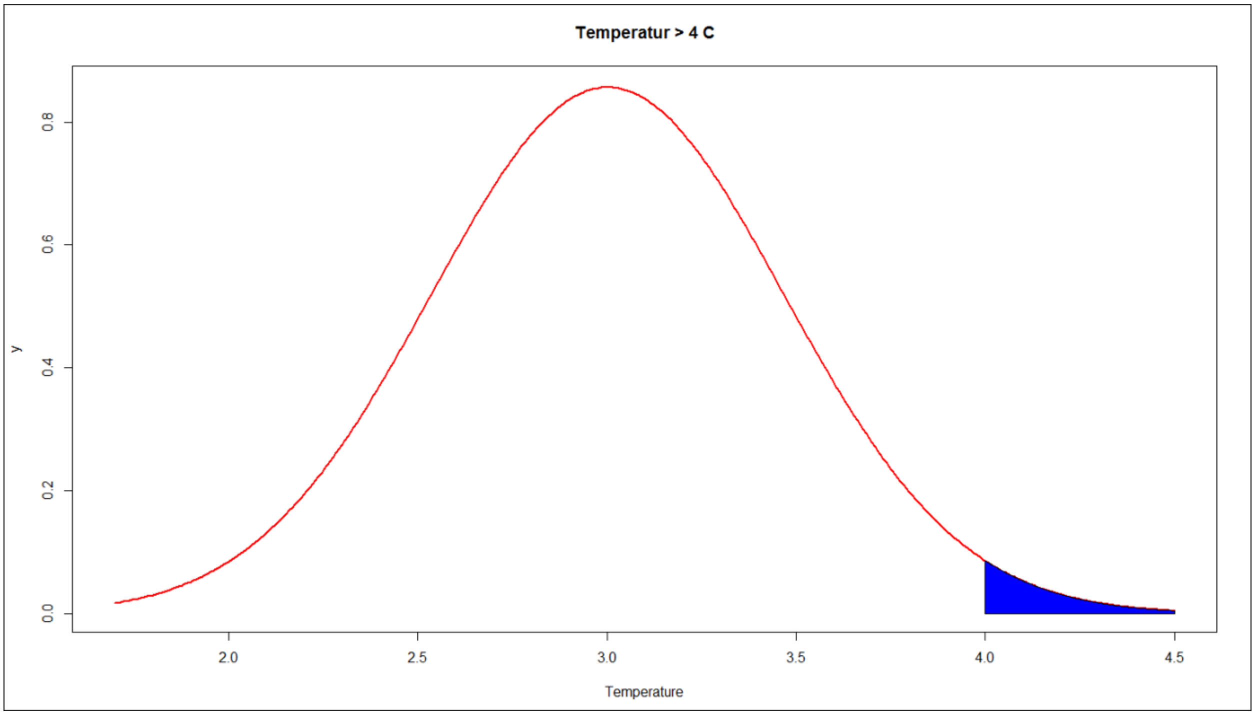
The value obtained is:

```
> pnorm(X)
[1] 0.01569904
```

5. Assuming normal distribution, find the probability of the temperature above 4ºC.

<u>Solution:</u>

To find: Probability (Temperature>4ºC)



By default, in R Studio, the area of the curve to the left of the Z value is taken for probability calculation. To find the area of the curve to the right of the Z value, "lower.tail = FALSE" should be mentioned in the "pnorm" command.

```
Y <- (4-3.002466)/0.4658319
print(Y)
pnorm(Y, lower.tail = FALSE)
```

The value obtained is:

```
> pnorm(Y, lower.tail = FALSE)
[1] 0.01612076
```

6. What will be the penalty for the AMC Company?

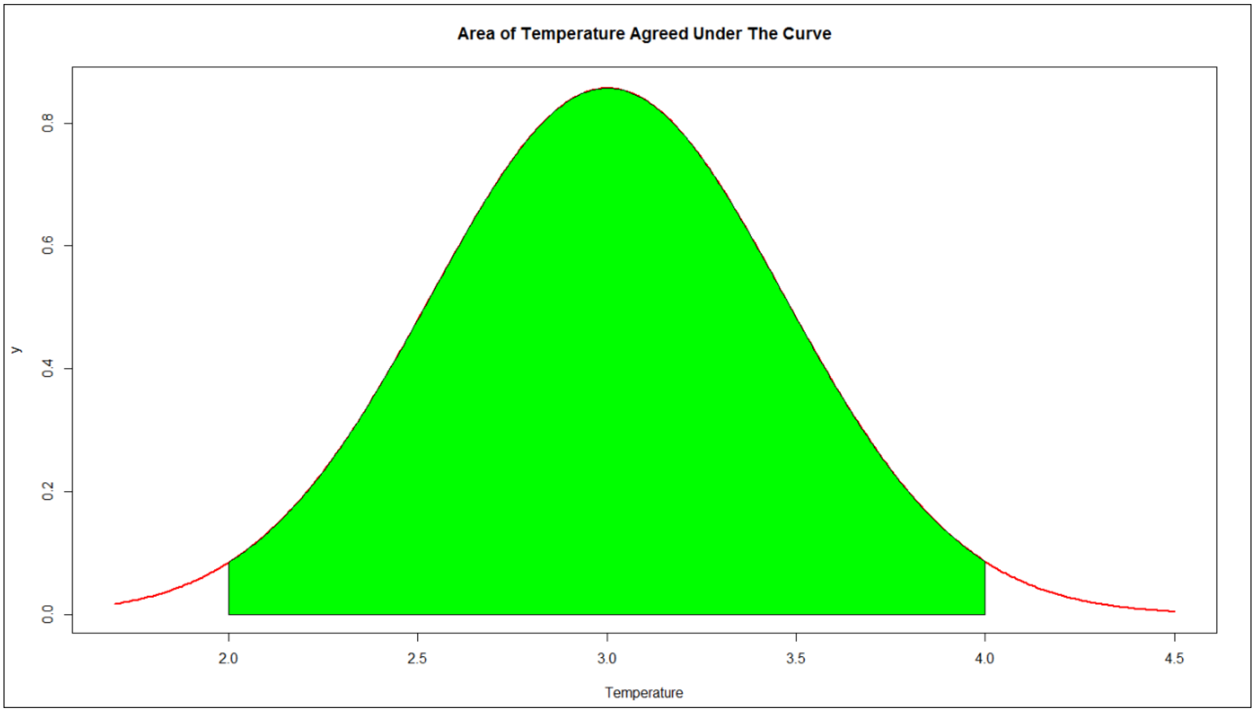
<u>Solution:</u>

The requirement by the company is that, the temperature range in the cold storage should be from 2ºC to 4ºC. But, on examining the dataset, we found that the temperature range in the cold storage is ranging from 1.7ºC to 4.5ºC.

```
> summary(df_coldstg$Temperature)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.700   2.700   3.000   3.002   3.300   4.500 
```

Hence, we have to find the percentage of the variation based on which the fine will be levied.

The area under the curve in between the temperatures 2ºC and 4ºC is:



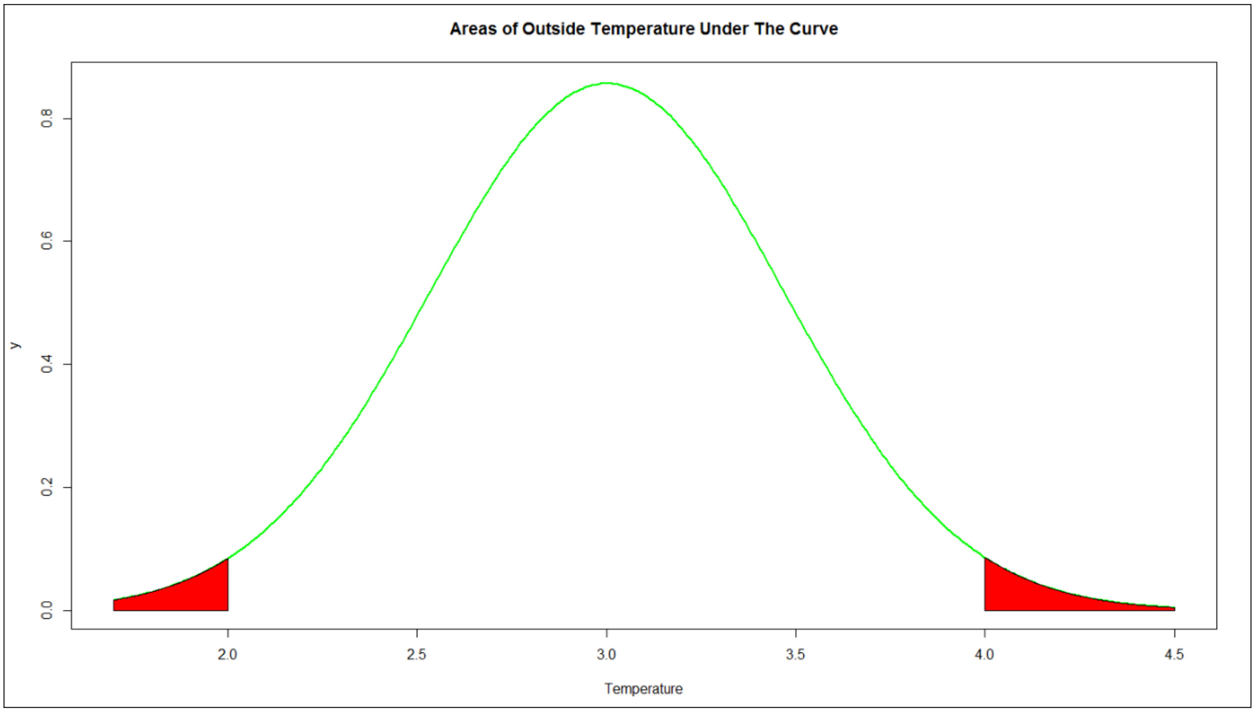

```
X <- (2-3.002466)/0.4658319
pnorm(X)
```

```
Y <- (4-3.002466)/0.4658319
print(Y)
pnorm(Y, lower.tail = FALSE)
```

```
A <- pnorm(Y)-pnorm(X)
print(A)
```

```
> print(A)
[1] 0.9681802
```

Similarly, the area under the curve in between the temperatures 1.7ºC and 4.5ºC is:




```
Z <- pnorm((4.5-3.002466)/0.4653819)-pnorm((1.7-3.002466)/0.4653819)
print(Z)
```

```
> print(Z)
[1] 0.9967888
```

The percentage variation in between the actual range and the obtained range:

```
> Z-A
[1] 0.02860863
```

Which is 2.8%.

The variation lies in the 2.5%-5% variation bracket.

<u>Inference:</u>

Since the variation between the specified range and the actual range is 2.8%, the penalty would be 10% of AMC (Annual Maintenance Cost).

<u>Problem 2:</u>

On checking the dataset given for the 35 days in the months of February and March, we assume that the values are normally distributed. Also, the supervisor demands that the temperature maintained should be less than or equal to 3.9ºC. Since the alpha (α) value is given as 0.1, the confidence level taken is 90%. Also, the summary and the standard of the values in the datasheet are as follows.

```
summary(df_march$Temperature)
```

```
> summary(df_march$Temperature)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.800   3.900   3.900   3.974   4.100   4.600 
```

The standard deviation of the temperature values are:

```
sd(df_march$Temperature)
```

```
> sd(df_march$Temperature)
[1] 0.159674
```

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

   Answer:

   "t-Test" for hypothesis is suitable to check whether corrective action needs to be taken or not in the cold storage as the population SD (S) is not known and assuming the population is normally distributed. Also, the values are continuous as it varies day to day.
2. State the Hypothesis, perform hypothesis test and determine p-value.

   Answer:

   - On checking the datasheet, we are able to find the following statistical values.

     X^{bar}(sample mean obtained from the dataset) =3.974
     μ (actual mean from the hypothesis) =3.9
     n=35
     S (population sd) = 0.159674
     df (degrees of freedom) =35-1=34

   - Also, the statements for the hypothesis are:

     H_o (null hypothesis) is **μ <= 3.9**
     H_1 (alternate hypothesis) is **μ > 3.9**

   - With the given values, we are able to find **T_{stat}** value using the below formula:

T_{stat}=(X^{bar}-µ)/(S/sqrt(n))

Solution:

```
tstat <- (3.974-3.9)/(0.159674/sqrt(35))
print(tstat)
```

```
> print(tstat)
[1] 2.741773
```

The p-value can be calculated using the T_{stat} value by using "pt" command.

```
pt(tstat, 34, lower.tail = FALSE)
```

```
> pt(tstat, 34, lower.tail = FALSE)
[1] 0.004836796
```

- At 90% confidence, the p-value obtained is **1.28**.

```
qnorm(0.9)
```

```
> qnorm(0.9)
[1] 1.281552
```

- On plotting the $T_{critical}$ value in the standard normal curve, the area of the curve beyond lies in the rejection zone.
- On checking, the alpha (α) value is 0.1 which is greater than the p-value which is 0.004836796.
- Hence, the null hypothesis is failed to accept.

3. Give your inference.

Inference:

On performing the hypothesis test, we can infer that the corrective action has to be taken since the temperature maintained in the cold storage is exceeding 3.9°C.