Project

Statistical Method of Decision Making

Haresh P Tayade
PGP-DSBA
Online Sept’2021

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 2o - 4o 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 the probability of temperature going outside the 2o - 4o 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 the date level is given in the file

“Cold_Storage_Temp_Data_.csv”.

Questions:

1. Find mean cold storage temperature for Summer, Winter, and Rainy Season. (7 marks)

2. Find the overall mean for the full year. (7 marks)

3. Find Standard Deviation for the full year. (7 marks)

4. Assume Normal distribution, what is the probability of temperature having fallen below 2o C? (7 marks)

5. Assume Normal distribution, what is the probability of temperature having gone above 4o C? (7 marks)

6. What will be the penalty for the AMC Company? (5 marks)

Find mean cold storage temperature for Summer, Winter, and Rainy Season. (7 marks)

```
df.head()
```

	Season	Month	Date	Temperature
0	Winter	Jan	1	2.3
1	Winter	Jan	2	2.2
2	Winter	Jan	3	2.4
3	Winter	Jan	4	2.8
4	Winter	Jan	5	2.5

	Season	Month	Date	Temperature
360	Winter	Dec	27	2.7
361	Winter	Dec	28	2.3
362	Winter	Dec	29	2.6
363	Winter	Dec	30	2.3
364	Winter	Dec	31	2.9

Head and Tail of DataFrame

Shape ,info,datatypes of dataframe

```
df.shape
```

```
(365, 4)
```

```
df.isna().sum()
```

```
Season         0
Month          0
Date           0
Temperature    0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 365 entries, 0 to 364
Data columns (total 4 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Season       365 non-null    object 
 1   Month        365 non-null    object 
 2   Date         365 non-null    int64  
 3   Temperature  365 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 11.5+ KB
```

```
df.dtypes
```

```
Season          object
Month           object
Date             int64
Temperature    float64
dtype: object
```

Description of Dataframe

```
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Date	365.0	15.720548	8.808321	1.0	8.0	16.0	23.0	31.0
Temperature	365.0	3.002466	0.465832	1.7	2.7	3.0	3.3	4.5

Checking for duplicates

```
df.duplicated().sum()
```

```
0
```

Observation

- **The above dataset contains 365 rows & 4 columns.**
  - **The variable month is season and month is object datatype and date is integer datatype , temperature is float datatype.**
  - **As we check they all are non null datatype.**
  - **The summary table shows mean, standard deviation, minimum & maximum values, etc. for all the variables.**

Question 1

Find mean cold storage temperature for Summer, Winter, and Rainy Season. (7 marks)

ANSWER

The mean cold storage temperature as per season wise as follows

```
Rainy = 3.0877049180327867
Summer = 3.1475
Winter = 2.7764227642276422
```

Summer **=** 3.1475

Winter **=** 2.7764227642276422

Rainy Season **=** 3.0877049180327867

Question 2

Find the overall mean for the full year. (7 marks).

Answer:

As before we checked the mean temp for summer ,winter, rainy season now we need to check the overall mean for the full year.

```
Overall mean of Temperature for the full year is 3.0024657534246546
```

Question 3

Find Standard Deviation for the full year. (7 marks).

Answer:

The overall standard deviation for the full year is

```
Standard Deviation of Temperature for the full year is 0.4658319416510761
```

Standard deviation of temperature for the full year is

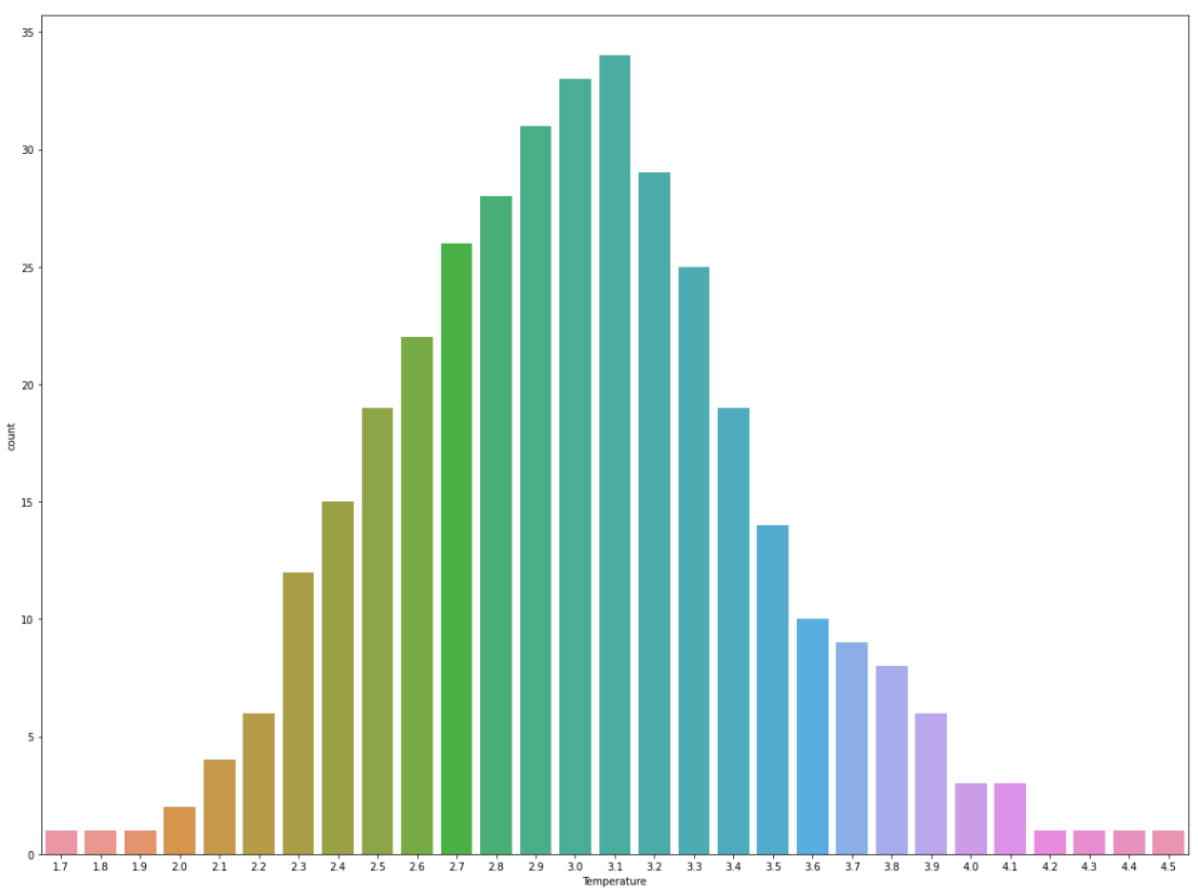
0.4658319416510761

Question 4

Assume Normal distribution, what is the probability of temperature having fallen below 2º C?

Answer:

Normal Distribution



Standard Deviation of Temperature for the full year is 0.4658319416510761.

Overall mean of Temperature for the full year is 3.0024657534246546.

Looking at the quest we need to find the probability of Temp falling below 2 Degree Celcius.

putting it into formula z = x - Mu /@. (Mu = Mean, @ = Standard Daviation)

z = (2 - 3.0024657534246546) / 0.4658319416510761

z = -2.151990157376403

After converting into whole number the calculation will be

Z = -2.0

Taking z = -2.151990157376403 the cdf will be

```
print(stats.norm.cdf(-2.151990157376403))
```

```
0.015699064791364483
```

This denotes that 1% is the probability of temp having fallen below 2 degree Celsius.

Taking z = -2.00 the cdf will be

```
print(stats.norm.cdf(-2.0))
```

```
0.022750131948179195
```

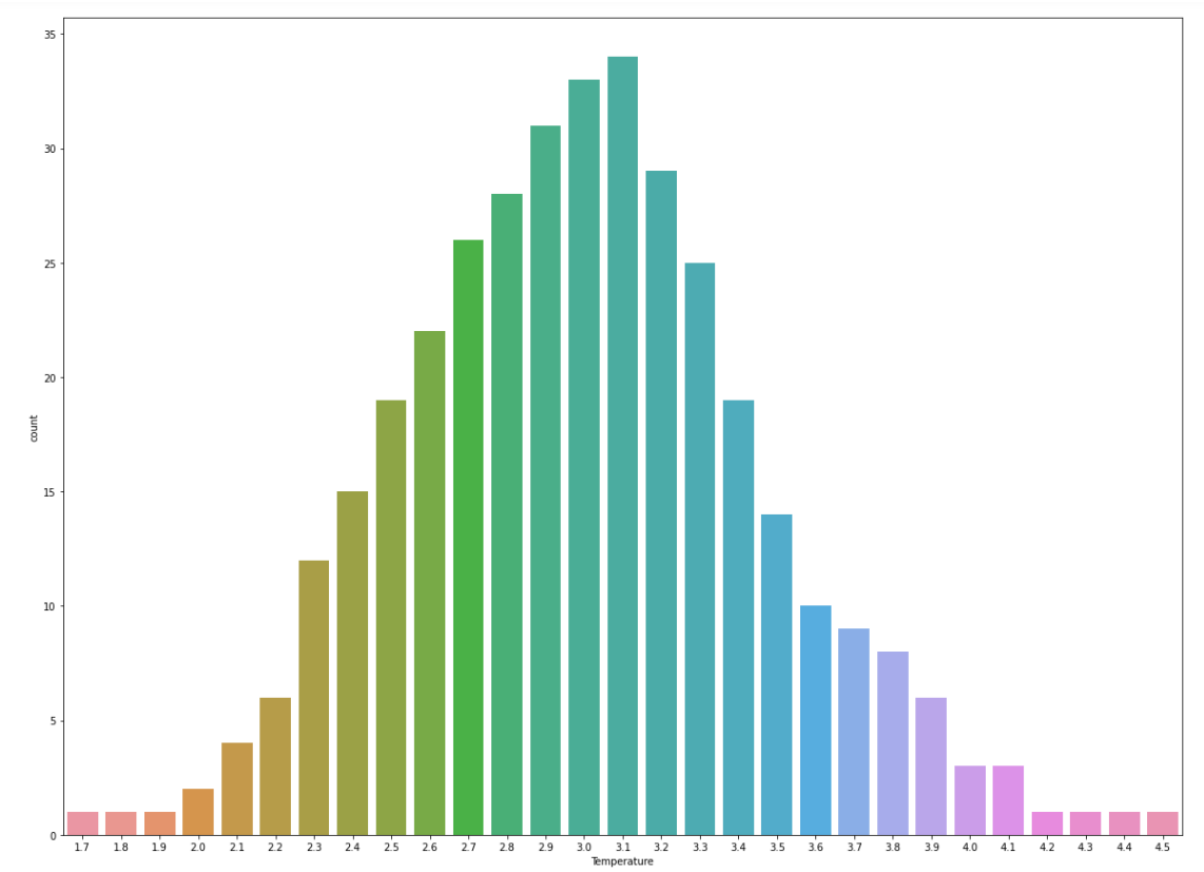
This denotes that 2% is the probability of temp having fallen below 2 degree Celsius.

Question 5

Assume Normal distribution, what is the probability of temperature having gone above 4º C? (7 marks)

Answer

Normal Distribution



Standard Deviation of Temperature for the full year is 0.4658319416510761.

Overall mean of Temperature for the full year is 3.0024657534246546.

Looking at the quest we need to find the probability of Temp falling below 4 Degree Celsius.

putting it into formula z = x - Mu /@. (Mu = Mean, @ = Standard Deviation)

z = (4 - 3.0024657534246546) / 0.4658319416510761

```
z = 2.146236559139785
```

z = 2.2

After converting into whole number the calculation will be

Z = 2.0

Taking z = 2.146236559139785 the cdf will be

```
1 - stats.norm.cdf(2.146236559139785)
```

```
0.015927055086573216
```

This denotes that 1% is the probability of temp having fallen below 4 degree Celsius.

Taking z = 2.00 the cdf will be

```
1 - stats.norm.cdf(2.0)
```

```
0.02275013194817921
```

This denotes that 2% is the probability of temp having fallen below 4 degree Celsius.

When Added the Z score 2.0 in the formula with -1 (as it is falling on the right side of the mean and has a positive output) in Python got the result as 0.0227501319481792 1. This denotes that 2% is the probability of temperature having fallen below 2º C

Question 5

What will be the penalty for the AMC Company? (5 marks)

Answer:

It was agreed that if it was statistically proven that the probability of temperature going outside the 2 degree – 4 degree Celsius 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.
Hence Looking at the Statistical Proven Data in the above case Study, we can infer that the Company would not be paying any Penalty. As it is clearly seen that the probability has not exceeded the 2% mark of the probability.

Problem 2

Hypothesis testing

Executive Summary

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 the last 35 days' temperatures. As a safety measure, the Supervisor has been vigilant to maintain the mean temperature 3.9º C or below. Assume 3.9º C as the upper acceptable 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 if corrective action is needed at the cold storage plant? Justify your answer. (6 marks)
2. State the Hypothesis and do the necessary calculations to accept or reject the corresponding null hypothesis. (8 marks)
3. Give your inference. (6 marks) Dataset to be used for Problem 2: Cold_Storage_Mar2018_.csv

Head and tail of dataset

	Season	Month	Date	Temperature
0	Summer	Feb	11	4.0
1	Summer	Feb	12	3.9
2	Summer	Feb	13	3.9
3	Summer	Feb	14	4.0
4	Summer	Feb	15	3.8

```
Cold.tail()
```

	Season	Month	Date	Temperature
30	Summer	Mar	13	4.2
31	Summer	Mar	14	4.2
32	Summer	Mar	15	3.8
33	Summer	Mar	16	3.9
34	Summer	Mar	17	3.9

Info and shape and null values and description of dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Season         35 non-null     object
 1   Month          35 non-null     object
 2   Date           35 non-null     int64
 3    Temperature   35 non-null     float64
dtypes: float64(1), int64(1), object(2)
memory usage: 1.2+ KB
```

Shape N Description of dataset

```
Cold.shape
```

```
(35, 4)
```

```
Cold.describe()
```

	Date	Temperature
count	35.000000	35.000000
mean	14.400000	3.974286
std	7.389181	0.159674
min	1.000000	3.800000
25%	9.500000	3.900000
50%	14.000000	3.900000
75%	19.500000	4.100000
max	28.000000	4.600000

Observation

- **The above dataset contains 35 rows & 4 columns.**
  - **The variable month is season and month is object datatype and date is integer datatype , temperature is float datatype.**
  - **As we check they all are non null datatype.**

- **The summary table shows mean, standard deviation, minimum & maximum values, etc. for all the variables.**

Question 1.

Which Hypothesis test shall be performed to check if corrective action is needed at the cold storage plant? Justify your answer. (6 marks)

Answer:

Test = 1 sample t test Because we do not have the Standard Deviation provided.

Level of Significance alpha = 0.1.

STD not given hence T - Test

n = 35
Test to be used is '1 sample t test'.

Reasoning: - We do not have the Standard Deviation provided in the problem. We only have the Level of Significance alpha = 0.1. Also, the Standard Deviation is not given, hence chose to perform or use '1 sample t test'.

Degree of Freedom: Since the sample is the same for both Sampling tests, we have N-1 degrees of freedom : 34.

Since the sole purpose of the test is to check whether the Temperature was 3.9 Degree or more, we would prefer a 1 sample t test.

Question 2

State the Hypothesis and do the necessary calculations to accept or reject the corresponding null hypothesis. (8 marks)

Answer:

H0: The Mean Temperature was equal to 3.9 Degree.

H1: The Mean Temperature was not equal and more than 3.9 Degree.

```
mn = Cold.Temperature.mean()
mn
```

```
3.974285714285715
```

```
t_statistic, p_value = ttest_1samp(Cold['Temperature'],3.9)
print('One sample t test \nt statistics: {0} P Values {1}'.format(t_statistic, p_value))
```

```
One sample t test
t statistics: 2.752358609800241 P Values 0.009422395404264431
```

Per the Query, assumed 3.9° C as the upper acceptable mean temperature and at alpha = 0.1.

As the P Value is less than alpha = 1. We have failed to Reject the Null Hypothesis.

It can be seen in the picture above that the Mean that was calculated was 3.974285714285715. Used the same in the formula and got the below results.

Per the Query, assumed 3.9º C as the upper acceptable mean temperature and at alpha = 0.1.

As the P Value is less than alpha = 1. We have failed to Reject the Null Hypothesis with the confidence level of 99%.

Question 3

Give your inference (6 marks)

Answer:

We have performed a Hypothesis Test to solve the given problem. As the Standard Deviation was not provided, we have chosen to use the '1 sample t test'.

Considering the test results, we can infer with the 99% confidence Level that the Mean Temperature was equal to 3.9 Degrees in those days given in the Sample data.

The End