Mini Project | Statistical Learning

Greatlearning

Problem1 - K2 Cold Storage Case-Study

K2 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 degrees.

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 the it was statistically proven that probability of temperature going outside the 2 - 4 degrees during the one-year contract was above 2.5% and less than 5% then the penalty would be 10% of AMC amount. 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 "K2_Cold_Storage_Temp_Data.csv"

1.Find mean cold storage temperature for Summer, Winter and Rainy Season?
2.Find overall mean for the full year.
3.Find Standard Deviation for the full year.
4.Assume Normal Distribution, what is the probability of temperature having fallen below 2 degree?'
5.Assume Normal Distribution, what is the probability of temperature having gone above 4 degree? (Optional)
6.What will be the penalty for the AMC Company?

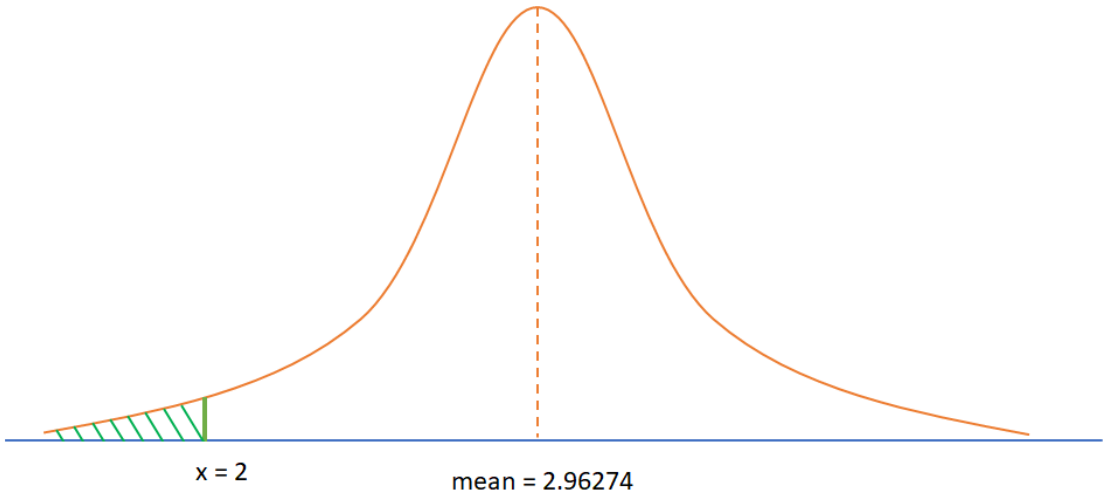
```
# set path here
# setwd(<path>)
Data_P1 <- read.csv("K2_Cold_Storage_Temp_Data.csv")
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Q1, Q2 and Q3

```
avg_winter_temp <- mean(Data_P1$Temperature[which(Data_P1$Season ==
"Winter")], na.rm = TRUE)
avg_summer_temp <- mean(Data_P1$Temperature[which(Data_P1$Season ==
"Summer")], na.rm = TRUE)
avg_rainy_temp <- mean(Data_P1$Temperature[which(Data_P1$Season == "Rainy")],
na.rm = TRUE)

mean_full_year_temp <- mean(Data_P1$Temperature)
sd_temp <- sd(Data_P1$Temperature)

paste("Average cold storage temperature during winter is ", avg_winter_temp)

## [1] "Average cold storage temperature during winter is  2.70081300813008"

paste("Average cold storage temperature during summer is ", avg_summer_temp)

## [1] "Average cold storage temperature during summer is  3.15333333333333"

paste("Average cold storage temperature during rainy is ", avg_rainy_temp)

## [1] "Average cold storage temperature during rainy is  3.03934426229508"

paste("Overall average temperature is ", mean_full_year_temp)

## [1] "Overall average temperature is  2.9627397260274"

paste("Standard deviation is ", sd_temp)

## [1] "Standard deviation is  0.508589031488563"
```

Q4: Probability of temperature having fallen below 2 degree is the probability of shaded region

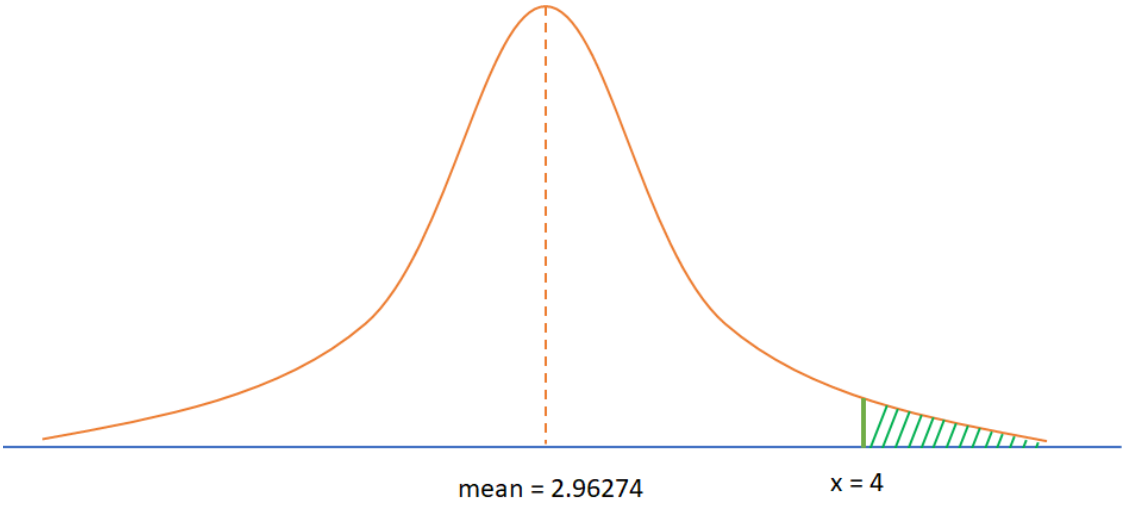
```
min_temp = 2
p_below_2  <- pnorm(q = min_temp,
                    mean = mean_full_year_temp,
                    sd = sd_temp,
                    lower.tail = TRUE)
print(paste("Probability of temperature having fallen below 2 degree = ",
p_below_2))

## [1] "Probability of temperature having fallen below 2 degree =
0.0291814594808524"
```



Q5 : probability of temperature having gone above 4 degree is the probability of shaded region

```
max_temp = 4
p_above_4 <- 1 - pnorm(q = max_temp,
                       mean = mean_full_year_temp,
                       sd = sd_temp,
                       lower.tail = TRUE)

# Below code can also be used
# p_above_4 <- pnorm(q = max_temp,
#                        mean = mean_full_year_temp,
#                        sd = sd_temp,
#                        lower.tail = FALSE)


print(paste("Probability of temperature having fallen above 4 degree = ",
p_above_4))
```

```
## [1] "Probability of temperature having fallen above 4 degree =
0.0207007672126939"
```




```
p = round((p_below_2 + p_above_4)*100,2)
print(paste("Total probability = ", p , "%"))

## [1] "Total probability =  4.99 %"
```

Since total probability, p = 4.99% which is less that 5%, penalty would be 10% of AMC amount.

Answers

1.Find mean cold storage temperature for Summer, Winter and Rainy Season?
#Winter mean = 2.70081
#Summer mean = 3.15333
#Rainy season mean = 3.039344
2.Find overall mean for the full year.
Overall mean = 2.96273
3.Find Standard Deviation for the full year.
Standard deviation = 0.5085
4.Assume Normal Distribution, what is the probability of temperature having fallen below 2 degree?' # P_below_2 = 0.02918
5.Assume Normal Distribution, what is the probability of temperature having gone above 4 degree? (Optional)
P_above_4 = 0.02070
6.What will be the penalty for the AMC Company?

Problem 2 - K2 Cold Storage Case-Study

In Mar 2018, K2 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 has been vigilant to maintain the temperature below 3.9 degree

Assume 3.9 degree as upper acceptable temperature range and at alpha = 0.1 do you feel that there is need for some corrective action in the Cold Storage Plant or is it that the problem is from procurement side from where K2 Cold Storage is getting the Dairy Products.

The data of the last 35 days is in "K2_Cold_Storage_Mar2018.csv"

1. State the Hypothesis, do the calculation using z test and give your inference.

2. State the Hypothesis, do the calculation using t test and give your inference. - (Optional)

```
Data_P2 <- read.csv("K2_Cold_Storage_Mar2018.csv")
```

Null Hypothesis and Alternate hypothesis

```
* H0: Temperature is below 3.9 degree, or T <= 3.9. There is no issue with
cold storage.
* H1: Temperature is above 3.9 degree, or T > 3.9. There is problem with cold
storage.

mean_temp = mean(Data_P2$Temperature)
sd_temp = sd(Data_P2$Temperature)
n = nrow(Data_P2)
x_bar = 3.9
```

Average temperature of cold storage = 3.97
Acceptable temperature(x_bar) = 3.9
Based on sample we can say that average temperature of the cold storage is greater than acceptable temperature. But stastical test is required to prove our observation.

Q1

```
z = (x_bar - mean_temp)/(sd_temp/sqrt(n))
probability = pnorm(z)
print(paste("p_value = ", probability))

## [1] "p_value =  0.00295838449736235"
```

Since p = 0.0029, is less than alpha (0.1), We'll reject the NULL hypothesis. With 90% confidence we can say that, **there is problem with cold storage** and there is need for some corrective action in the Cold Storage Plant

Q2

```
t = (x_bar - mean_temp)/(sd_temp/sqrt(n))
probability = pt(t, n-1)
print(paste("p_value = ", probability))

## [1] "p_value =  0.0047111977021326"
```

Since p = 0.0047, is less than alpha (0.1), We'll reject the NULL hypothesis. With 90% confidence we can say that, **there is problem with cold storage** and there is need for some corrective action in the Cold Storage Plant.

Alternatively we can use t.test to reject/fail to reject NULL hypothesis.

```
t.test(Data_P2$Temperature, mu = mean_temp, alternative = "greater",
conf.level = 0.1)

##
##  One Sample t-test
##
## data:  Data_P2$Temperature
## t = 0, df = 34, p-value = 0.5
## alternative hypothesis: true mean is greater than 3.974286
## 10 percent confidence interval:
##  4.00956     Inf
## sample estimates:
## mean of x
##  3.974286
```

Rejection region