Statistical Analysis, Simulation, and Modeling.
Mid-term project assignment
(Experimentation of Comparing Weather Data with Statistical Testing)

Cho Keunhee 26001904131
College of Information Science and Engineering, Ritsumeikan University, Japan

1. Introduction

The purpose of this study is to analyze the average temperature values using statistical approaches by utilizing the weather data from Seoul and Kusatsu. It is intended to conduct experiments to verify the following three hypotheses.

I. There is no difference, on average, in air temperatures registered in Seoul and the Kusatsu city in August 2019.
II. On average, October was cooler in 2020 than in 2018 in Seoul.
III. On average, there is no difference between average air temperature in December 2018 and 2019 in the Kusatsu city area.

2. Experimental design

For the first hypothesis, data sets are independent data of two different location. Therefore, this study will conduct an independent two sample z-test.
For the second hypothesis, data sets are paired data of same location with different year. Moreover, the hypothesis is only interested at one-side. Therefore, this study will conduct a Right-tailed, paired-sample z-test.
For the third hypothesis, data sets are paired data of same location with different year. Moreover, is interested in both tails. Therefore, this study will conduct a paired-sample z-test.

3. Data collection

The data used in this study was extracted from Korea and Japan's official Meteorological Administration. For the convenience of analysis, the extracted data is stored in a csv file. The following is the box-whisker plot of the data.
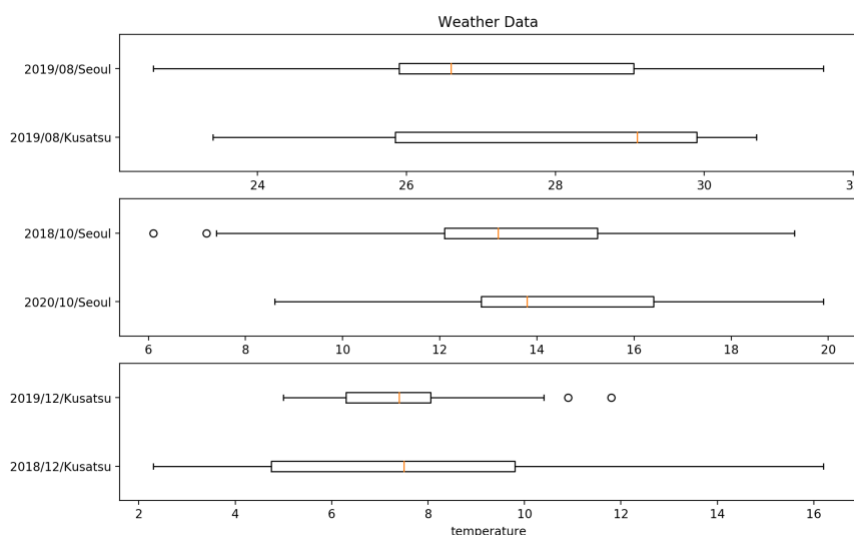
4-1. Data analysis (Q1)
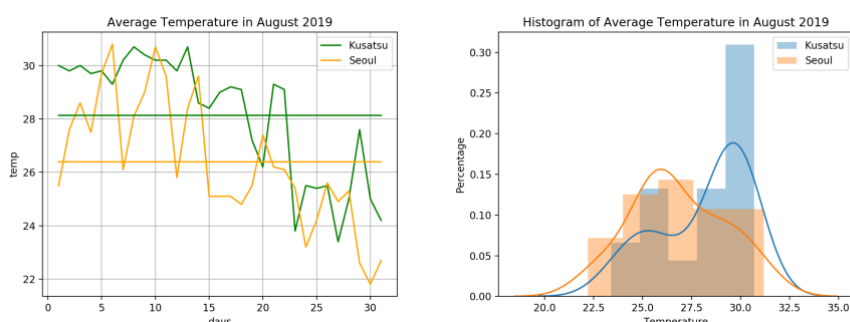
H_0: The average temperature of Kusatsu and Seoul is the same.
($\mu_1 = \mu_2$)
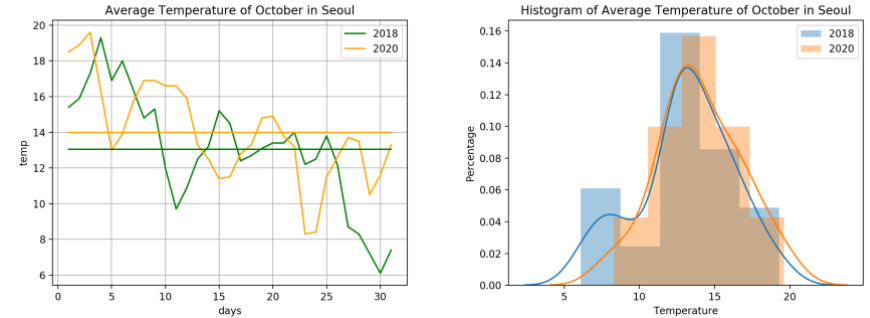H_1: The average temperature of Kusatsu and Seoul is different.
($\mu_1 \neq \mu_2$)
Independent two sample z-test: $z \approx -2.2730 \notin [-1.9599, +1.9599]$
Therefore, reject H_0 and take H_1
p-value $2*(1 - \Phi(z)) \approx 0.0230$

4-2. Data analysis (Q2)

H_0: The average temperature of Oct 2020 is not cooler or than 2018.
($\delta \leq 0$)
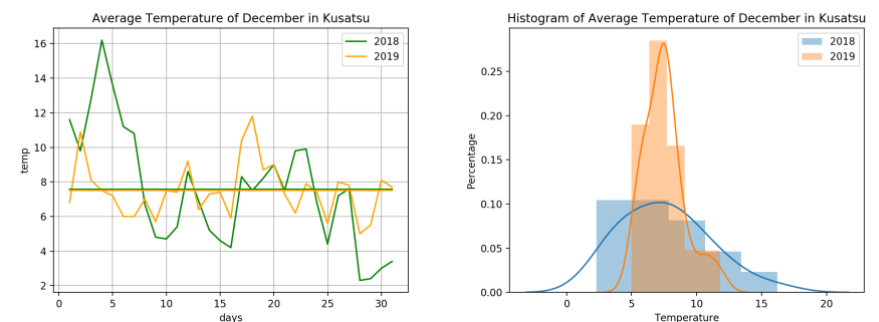H_1: The average temperature of Oct 2020 is cooler than 2018.
($\delta > 0$)
Paired-sample z-test: $z \approx -1.5574 \in [-1.6448, +1.6448]$
Therefore, it is not possible to reject H_0
p-value $(1 - \Phi(z)) \approx 0.0596$

4-3. Data analysis (Q3)

H_0: The average temperature of Dec 2018 and 2019 is the same.
($\delta = 0$)
H_1: The average temperature of Dec 2018 and 2019 is different.
($\delta \neq 0$)
Paired-sample z-test: $z \approx 0.0899 \in [-1.9599, +1.9599]$
Therefore, it is not possible to reject H_0
p-value $2*(1 - \Phi(z)) \approx 0.9283$

5. Results

For Hypothesis 1, it was possible to reject the null hypothesis. This states that there is a temperature difference between Kusatsu city and Seoul in August 2019.

For Hypothesis 2, it was not possible the reject the null hypothesis. Moreover, considering the fact that p-value is greater than 0.05 there is a weak evidence towards the null hypothesis. This states that there is a weak evidence of support the statement of temperature of October 2020 is not cooler than October 2018.

For Hypothesis 3, it was not possible to reject the null hypothesis. Moreover, considering the fact that p-value is 0.93 there is a significant evidence towards the null hypothesis. This states that the average temperature of December 2018 and 2019 in Kusatsu is the same.

6. Conclusion

This study has analyzed the average temperature data of August, October, and December in Kusatsu city and Seoul. Although there was a weak evidence of average temperature of 2020 October in Seoul higher than 2018 October. However, there was no difference of average temperature in Kusatsu between 2018 and 2019 December. The conclusion is that it is difficult to accept arguments that external factors may have worked in common for both cities.