New Haven Temperatures

Alberta "Albi" Kovatcheva

Background

The 'nhtemp' built-in data set gives the mean annual temperature in New Haven, CT, for the years 1912 to 1971.

Problem

A researcher wants to test whether the average over the first 25 years (1912 to 1936) of the data is statistically significantly different than the average over the last 25 years (1947 to 1971) of the data set. The researcher created two vectors from the data set: first25 and last25, using the following code:
first25 ← nhtemp[1:25]
last25 ← nhtemp[36:60]

Task

The scientist must compute a test to see if these two vectors have the same mean.
To do this, they must first determine type of t-test to use: Single sample, Independent & Dependent t-tests.

t-Test

Because the data is paired, the appropriate statistical test in this situation is the dependent t-Test. The scientist is investigating temperature change between the first and last 25 years of data, which means that the data are related by time.

Assumptions

The underlying assumption behind this test is that both vectors, 'first25' and 'last25' have the same count, that the data is paired, and that the data is normally distributed.

Hypotheses

My null hypothesis (H0) is that the two means are the same, so that there is no difference between them.
On the other hand, my alternative hypothesis(H1) is that the two means are different.
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Results of the Hypothesis Test

The p-value measures the probability of finding a particular set of observations if the null hypothesis were true.

In this case, p= 0.0006383. This is less than 0.05, which is the significance level and probability of wrongly rejecting the null hypothesis. Because $p < \alpha = 0.05$, the researcher has adequate evidence to reject the null hypothesis and instead accept the alternative hypothesis. The results of the hypothesis show that there is a statistically significant difference between the temperatures during the first and last 25 years of data collection.

Graphical Interpretation of Dependent t-Test Results

A histogram of the difference of New Haven Temperatures during the first and last 25 years of data shows that the data is approximately normally distributed. The purple line represents the mean difference of the first and last 25 years of temperature measurements.

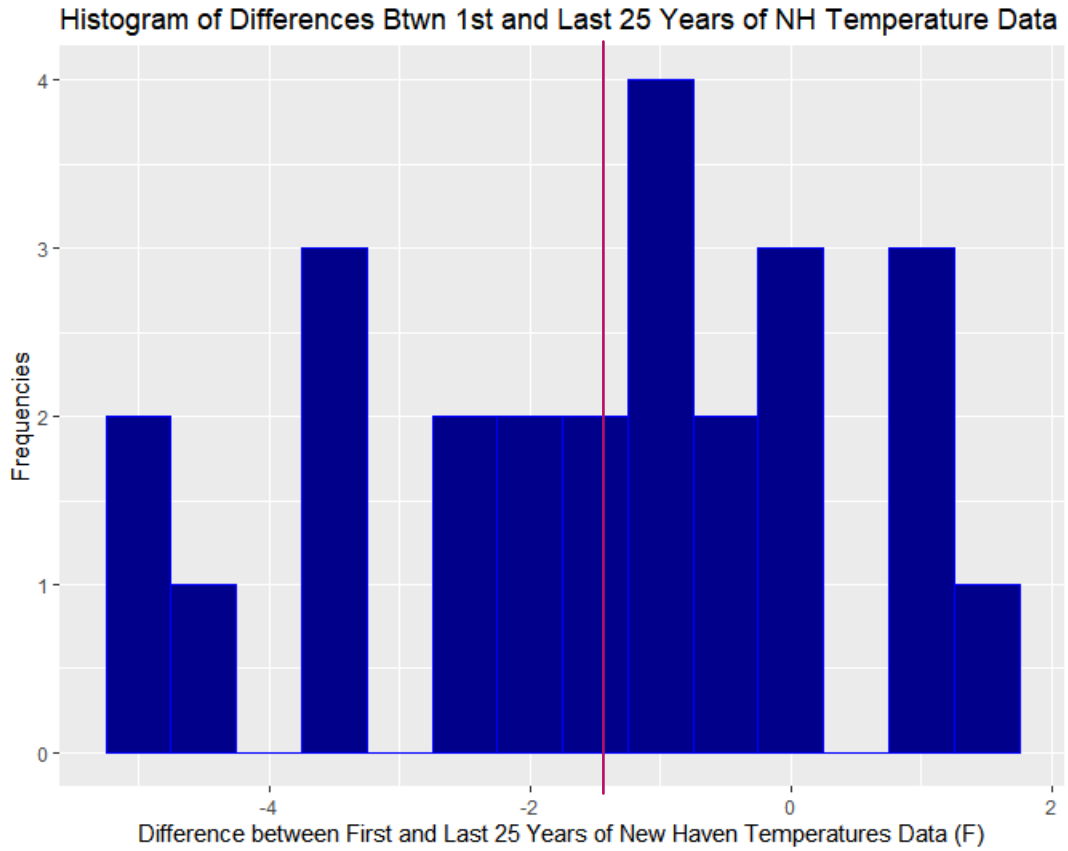

FIGURE 1: HISTOGRAM OF NEW HAVEN TEMPERATURE DIFFERENCES BETWEEN FIRST AND LAST 25 YEARS OF DATA

A Q-Q plot of the difference of New Haven Temperatures during the first and last 25 years of data further confirms that the data is approximately normally distributed.

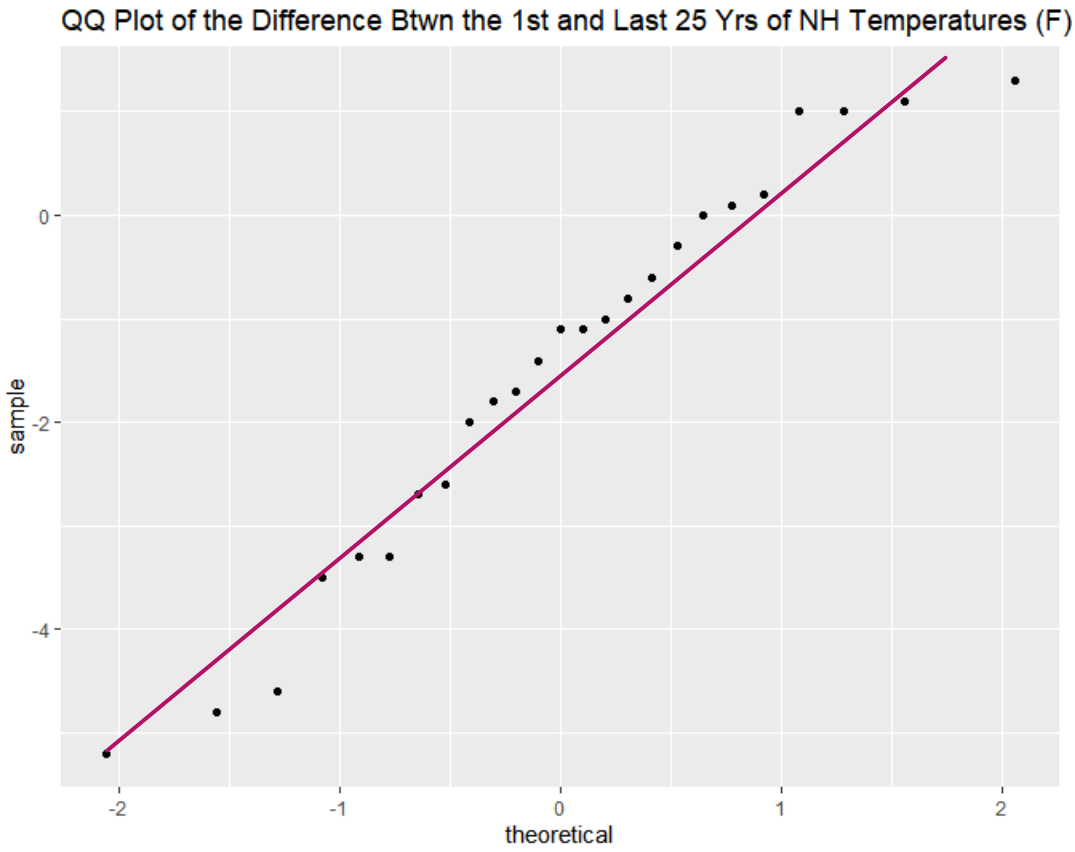

FIGURE 2: QQ PLOT OF THE DIFFERENCE BETWEEN THE 1ST AND LAST 25 YEARS OF NEW HAVEN TEMPERATURES DATA.

Conclusion

Because the data is paired and approximately normally distributed, it meets the underlying assumption for the dependent t-test. Therefore, the researcher can have confidence in their test results. In this situation, $\mu_1 = 50.416$ and $\mu_2 = 51.9$, and the mean difference between the first and last 25 years of New Haven temperature data is -1.484 Fahrenheit. This means that data shows an increase in temperatures between the first and last 25 years of data collection.