MAT 243 Final Project Part 2: Summary Report

Milestone II

Scenario (A): Title *National Weather Service*

Jason Lima

Jason.lima1@snhu.edu

Southern New Hampshire University

**1. Statement of the Problem**

The purpose of this analysis is to compare two hypotheses of data across the same data pool to determine if the observed values differ from each other and how. Weather can be different in different months and within the same month, temperature can be different in different years. Questions analyst need to ask is, “Is there a trend that occurs in a single month? And does this trend carry over to the next?” This would lead to investigating the cause and effect of this relationship, depending on the statistical significance of the correlation. To answer these types of questions, we will implement the workflow for two sample hypothesis tests for proportion and mean hypothesized differences.

**2. Hypothesis Test for Population Proportions (2-sample)**

Table 1. Hypothesis Test for EMXT Proportion

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Count of July Months | 29 |
| Count of August Months | 29 |
| Sum count of July values greater than 32.5 Celsius | 15 |
| Sum count of August values greater than 32.5 Celsius | 11 |
| Significance Level | 1% (0.01) |
| z-score | 1.0561 |
| P-value | 0.2909 |

In the scenario for Table 1, we are looking to find if the proportion of Extreme Maximum Temperature over 32.5 degrees Celsius is the same for months July and August. Given the results we are using a 1% significance level. There is the count of data point in each given month. Then the count of the months that meet the hypothesis criteria which is months with an EMXT greater than 32.5 degrees Celsius. The z-score represents the number of standard deviations from the mean. The positive z-score indicates that the quantity is above the mean. About one standard deviation above the mean. The p-value is the probability of obtaining a result that is as extreme or more extreme than the data if the null hypothesis were true. Since the p-value is well above the 1%, we fail to reject the null hypothesis.

Table 2. Hypothesis Test for EMXP Proportion

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Count of February Months | 29 |
| Count of August Months | 29 |
| Sum count of February values greater than 20.0mm | 13 |
| Sum count of August values greater than 20.0 mm | 21 |
| Significance Level | 5% (0.05) |
| z-score | -2.1328 |
| P-value | 0.0329 |

The counts for Table 2 are the same as Table 1. The difference between these tables is Table 2 is evaluating the difference of two population proportions for Extreme Maximum Precipitation over 20.0mm. Also, the population proportions will be evaluated using months of February and August. The z-score represents a value a little over 2 standard deviations below the mean of the populations. The level of signifincance for the hypothesis is 5% or 0.05. Since the p-value of 0.0329 is less than 0.05, there is enough evidence to reject the null hypothesis. The hypothesis is to check if the proportion of Extreme Maximum Precipitation of over 20.00mm is the same for February and August. The null hypothesis is the alternative, the proportion of EMXP above 20.0mm is not the same.

**3. Hypothesis Tests for Population Means (2-sample)**

Table 1. Hypothesis Test for EMXT Mean

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| jul\_data | An array of EMXT values where the ‘Month’ equals 7 |
| aug\_data | An array of EMXT values where the ‘Month’ equals 8 |
| Statistic | 1.2931 |
| P-Value | 0.2013 |

Using the previously used significance level of 1% (0.01), if we look at the p-value, we can analyze that since the p-value is greater then the significance level, there is insufficient evidence to reject the null hypothesis of the average EMXT is the same for July and August.

Table 1. Hypothesis Test for EMXP Mean

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| feb\_data | An array of EMXP values where the ‘Month’ equals 2 |
| aug\_data | An array of EMXP values where the ‘Month’ equals 8 |
| Statistic | 2.2438 |
| P-Value | 0.0299 |

Using the previously used significance level of 5% (0.05), since the p-value is less than the significance level, we have enough evidence to reject the null hypothesis of the EMXP in February is more than August.

**4. Results**

After performing a hypothesis test for the difference of two population proportions of Extreme Maximum Temperature, the results determine the z-score was 1.0561 standard deviations above the random normalized mean. Around 34.1% above the mean. The p-value, which gives evidence to accept or reject the hypothesis was 0.2909. With a provided statistical significance of 1% (0.01), since the p-value was above what was considered a statistical significance then we have enough evidence to reject the hypothesis. This means that the proportion of temperature above 32.5 degrees Celsius is not the same for months July and August. Now we can investigate factors that could have caused this hypothesis to not be true.

After conducting the analysis of the hypothesis test for the difference of two population proportions of Extreme Maximum Precipitation. The hypothesis goal was to determine if the proportion of precipitation above 20.0mm is the same for months February and August in the data set. Based on a z-value of -2.1328, means the comparison of the proportion is slightly greater than 2 standard deviation below the mean of the normally distributed dataset. Further, the p-value was below the 5% level of significance which determines the hypothesis had enough evidence to be accepted. We can look at factors that determine this hypothesis to be accepted such as the February and months and the August months experienced similar seasonal chances. One lead to cold weather and the other leads to warm weather.

**5. References**

Berrier, J., Nestler, S., Pardoe, I., Sturdivant, R. X., Watts, K., Chan, C., . . . Vahid, F. (2016). MAT 243: Applied Stats for STEM. Los Gatos, CA: Zyante.

Khan Academy. (n.d.). Retrieved from https://www.khanacademy.org/