Is the probability of a total eclipse directly correlated to the months in the year?

During the visual analysis above, it seemed that solar eclipses had a high count during the months of spring and summer and a low count during the months of fall and winter. But lunar eclipses seemed to be relatively the same count for all months.

For the statistical analysis, I have divided the data into two groups based on monthly analysis for both solar and lunar eclipses.

Group 1: September - February (Fall - Winter according to US seasons)

Group 2: March - August (Spring - Summer according to US season)

HYPOTHESIS TEST: To test the hypothesis that total eclipse is not directly correlated to the months in the year, we will assume the following null and alternative hypothesis. This hypothesis test will be performed for both solar and lunar eclipses.

Ho : The probability of total eclipse occurring is the same for Group 1 and Group 2. ($p1=p2$)

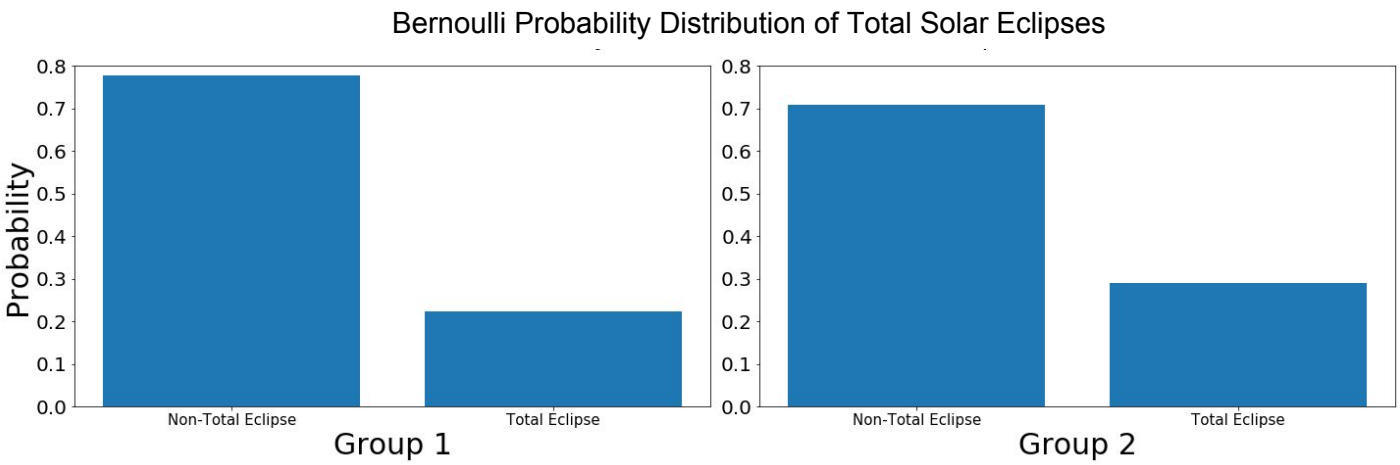
Ha : The probability of total eclipse occurring is not the same for Group 1 and Group 2.($p1 \neq p2$)

Hypothesis Test for Total Solar Eclipses

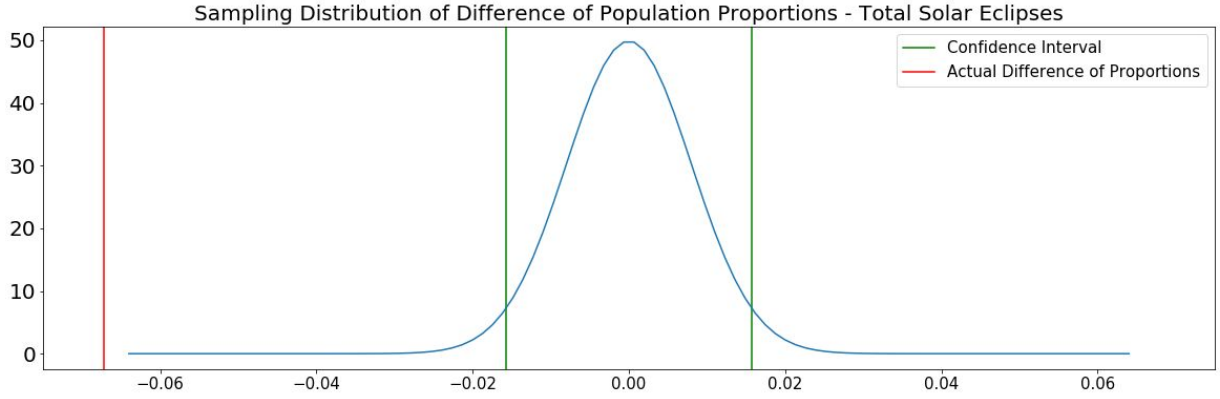
At first, I created dictionaries for each group, with key: [month,yearI and values of 1 (total eclipse) and 0 (non-total eclipse). Using these dictionaries, the probability that the data was a total eclipse in each group was calculated using:

$$\frac{sum\ of\ the\ dictionary\ values\ (adding\ all\ the\ 1s\ from\ total\ eclipses)}{length\ dictionary\ (all\ eclipses)}$$

The probabilities were then used to plot a Bernoulli probability distribution for visual analysis. The x-axis consists of Total Eclipse (p1 and p2) and Non-Total Eclipses (1-p1 and 1-p2).



To perform the statistical frequentist test, I assumed a 95% confidence interval. using the critical Z-value of 1.96 for a two-tail distribution. Using the calculated standard deviation difference between the two groups, the margin of error was calculated with z-value times the standard deviation difference: $z * \sigma_{p1-p2} \approx 0.01568$. Using the margin of error the confidence interval was calculated to be [-0.01568 0.01568] since the null hypothesis assumes that the mean of the distribution $\mu_{p_1-p_2} = p_1 - p_2 = 0$ since $p_1 = p_2$. The actual difference in proportions was calculated to be $\approx$ -0.06719, which is outside the confidence interval. Therefore, we can reject the null hypothesis which states that the total solar eclipses are not directly correlated to the months/seasons. All these calculations were plotted for better visualization.



Hypothesis Test for Total Lunar Eclipses

The same tests were performed on lunar eclipses. I assumed a 95% confidence interval. using the critical Z-value of 1.96 for a two-tail distribution. Using the calculated standard deviation difference between the two groups, the margin of error was calculated with z-value times the standard deviation difference: $z * \sigma_{p1-p2} \approx 0.01149$. Using the margin of error the confidence interval was calculated to be [-0.01149 0.01149] since the null hypothesis assumes that the mean of the distribution $\mu_{p_1-p_2} = p_1 - p_2 = 0$ since $p_1 = p_2$. The actual difference in proportions was calculated to be $\approx$ -0.00560, which is inside the confidence interval. Therefore, we cannot reject the null hypothesis which states that the total solar eclipses are not directly correlated to the months/seasons. All these calculations were plotted for better visualization.