Analyzing Rainfall for Urbana

[redacted]

University of Illinois, [redacted]

Abstract. *The objective of this laboratory is to determine the trends in spring and summer rainfall for Urbana. The Illinois State Climatologist and the Illinois State Water Survey supplied the rainfall data, which were then computed to determine the changes in the mean or variance in precipitation amount, trends in precipitation amount, changes in the mean or variance of peak rainfall and changes in the number of rain days. According to the t-test, there is no significant difference in the spring, and a significant difference in the summer for changes in the mean in precipitation. For the trends in precipitation amount, there is no trend line in the amount of precipitation throughout the years. In the spring, the precipitation amounts of 9, 10.5, and 12 inches occurred the most frequently at 18. In the summer, the precipitation amounts of 11 and 15 inches occurred the most frequently at 24 and 21 respectively. For the changes in the mean of peak rainfall, there is no significant difference in the data for spring and there is a significant difference in the data for summer. Finally, for the changes in the number of rain days, both spring and summer have a significant difference in their data. After determining the probability of non-exceedance, it was determined that the changes that any particular average precipitation amount will be exceeded are roughly the same as the chances of the 1901-2014-reference period of Urbana's rainfall data.*

Keywords. *Precipitation, Probability of Non-Exceedance, Mean, t-Test*

Introduction

The purpose of this exercise is to determine the spring and summer rainfall trends for Urbana. Given the amount of precipitation, the daily max precipitation, and the rain days from the year 1901 to year 2014, the computation of the mean and standard deviation allows the determination of changes in precipitation amount, peak rainfall, and number of rain days. By creating tables, graphs and histograms for precipitation, the trends for precipitation amount, frequency of precipitation amount, and the probability of exceedance can be determined.

Methodology

Using the spring and summer rainfall data for Urbana that were given, the information of the amount of precipitation, the amount of rain days, and the daily max precipitation from year 1901 to 2014 were displayed. The data was then parsed into Microsoft Excel to better organize the data. Since the directions asked for both spring and summer data analysis, the work was split in half, with one focusing on spring and the other focusing on summer.

Beginning with the daily max precipitation data, the creation of three columns, each being labeled “top”, “bottom”, and “diff” respectively. The purpose of the “top” column is to find the average of the daily max precipitation values between two consecutive years, such as 1901 and 1902. The function that was inputted into the cell that correlated with the second year, and by doing so, the average of the daily max precipitation can be determined between the two years. The function was then anchored to repeat the same function with the rest of the years. The purpose of the “bottom” column is the find the average of the daily max precipitation from the values of the years under the specified years. For example, for year 1902, the average was taken from years 1903-2014. The last cell, which was for year 2014, was renamed into “maxd”, which was used for the function to determine the average at a specific year. The cell was then anchored, allowing the values for the remaining years to be determined. The difference of between the “top” and “bottom” columns at each year was calculated, starting with the value 20 years from the top, and ending with value 20 years from the bottom. The data analysis tool was then used to find out the t-Test: Two-Sample Assuming Equal Variances values. The values that were used for the t-Test started with the highest number in the difference column and ended with the last number in the column. Next, the t-Test is calculated to determine the change in the means or variance of the daily max precipitation.

For changes in the number of rain days, the exact procedure is followed as in the daily max precipitation data. The only difference is that instead of using the daily max data, the data for the number of rain days is used.

Next, the trends in precipitation amount are determined. A scatterplot is created using the data for the amount of precipitation in each year. Once the points have been plotted, a trend line can be determined, which then determines the equation of precipitation in terms of a specific year. R^2 can also be determined from the trend line. The t-test is computed using the same steps as stated above. After the t-test is conducted, the frequency of the amount of precipitation is determined by using the data analysis tool, using the histogram option and the precipitation amount data. Once the points are determined, the creation of a histogram with frequency on the y-axis and precipitation in inches in the x-axis shows how often certain amounts of precipitation occur.

Finally, with distributions in precipitation amounts, only the year and the correlating precipitation amount values are needed. The columns are then shorted out by amount of precipitation from smallest to largest. Then, a new column is created, labeled “Rank”, and each year is ranked, with 1 being the lowest amount of precipitation and 113 being the highest. The probability of non-exceedance is calculated, using the equation, PNE=2*(Rank)/229. In order to find the probability of PNE (normal), the mean and standard deviation for precipitation amounts is needed. This can be determined by using functions on excel, such as average and stdev of series of values. The PNE (normal) for each year using the function for statistics called NORM.DIST, which determines the PNE (normal) using the data the amount of precipitation of the year, the mean, and the standard deviation. Once the first cell is determined, the data is then anchored for the remaining years. Two graphs are created on the same graph, the first with precipitation on the x-axis and PNE on the y-axis and the second with precipitation on the x-axis and PNE (normal) on the y-axis.

All the results were recorded.

Results

Table 1. **t-Test for Daily Max Precipitation for Spring**

	Variable 1	*Variable 2*
Mean	1.798139535	1.691594203
Variance	0.521910742	0.359054774
Observations	43	69
Pooled Variance	0.421236144	
Hypothesized Mean Difference	0	
df	110	
t Stat	0.844931253	
P(T<=t) one-tail	0.199991616	
t Critical one-tail	1.658824187	
P(T<=t) two-tail	0.399983233	
t Critical two-tail	1.981765282	

Table 2. **t-Test for Daily Max Precipitation for Summer**

	Variable 1	*Variable 2*
Mean	1.951608877	2.060037628
Variance	0.016448901	0.002783516
Observations	66	46
Pooled Variance	0.010858516	
Hypothesized Mean Difference	0	
df	110	
t Stat	-5.417527119	
P(T<=t) one-tail	1.79789E-07	
t Critical one-tail	1.658824187	
P(T<=t) two-tail	3.59578E-07	
t Critical two-tail	1.981765282	

Table 3. **t-Test for Rain Days for Spring**

	Variable 1	*Variable 2*
Mean	29.88461538	34.5862069
Variance	37.14615385	31.96631917
Observations	26	87
Pooled Variance	33.1329486	
Hypothesized Mean Difference	0	
df	111	
t Stat	-3.654449197	
P(T<=t) one-tail	0.000197617	
t Critical one-tail	1.658697265	
P(T<=t) two-tail	0.000395234	
t Critical two-tail	1.981566757	

Table 4. **t-Test for Rain Days for Summer**

	Variable 1	*Variable 2*
Mean	25.91703532	27.08797227
Variance	0.75134382	0.044986395
Observations	90	22
Pooled Variance	0.616493766	
Hypothesized Mean Difference	0	
df	110	
t Stat	-6.270363991	
P(T<=t) one-tail	3.62452E-09	
t Critical one-tail	1.658824187	
P(T<=t) two-tail	7.24904E-09	
t Critical two-tail	1.981765282	

Table 5. **t-Test for Precipitation for Spring**

	Variable 1	*Variable 2*
Mean	10.46135135	11.32842105
Variance	14.36311757	10.22253081
Observations	37	76
Pooled Variance	11.56542381	
Hypothesized Mean Difference	0	
df	111	
t Stat	-1.271866917	
P(T<=t) one-tail	0.10303934	
t Critical one-tail	1.658697265	
P(T<=t) two-tail	0.20607868	
t Critical two-tail	1.981566757	

Table 6. **t-Test for Precipitation for Summer**

	Variable 1	*Variable 2*
Mean	10.78071429	12.88651163
Variance	14.50847919	15.43650421
Observations	70	43
Pooled Variance	14.85962379	
Hypothesized Mean Difference	0	
df	111	
t Stat	-2.819399978	
P(T<=t) one-tail	0.002849678	
t Critical one-tail	1.658697265	
P(T<=t) two-tail	0.005699355	
t Critical two-tail	1.981566757	

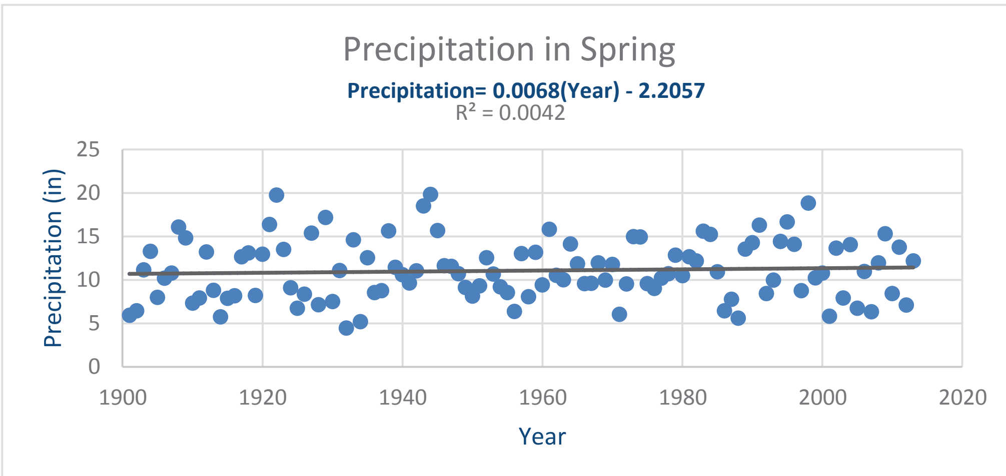

Figure 1. Amount of Precipitation throughout the Years in Spring

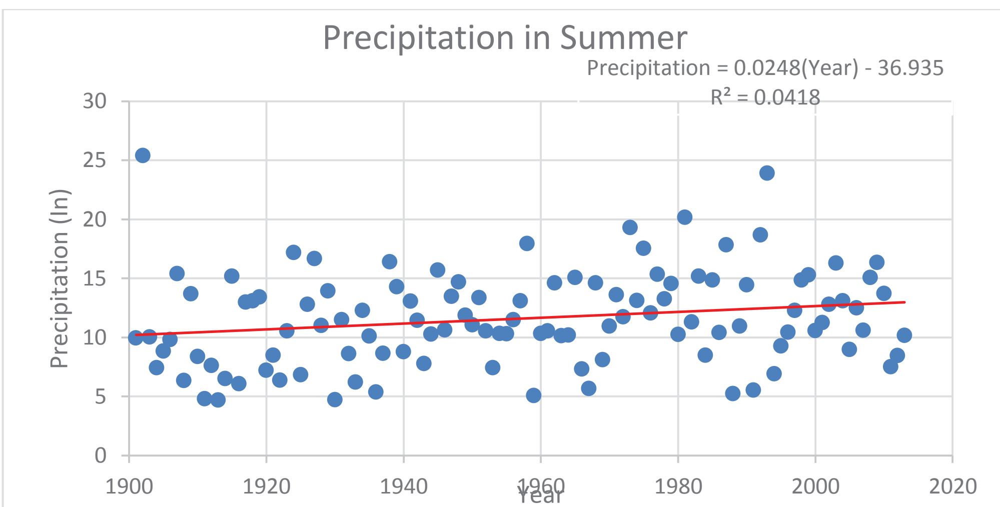

Figure 2. Amount of Precipitation throughout the Years in Summer

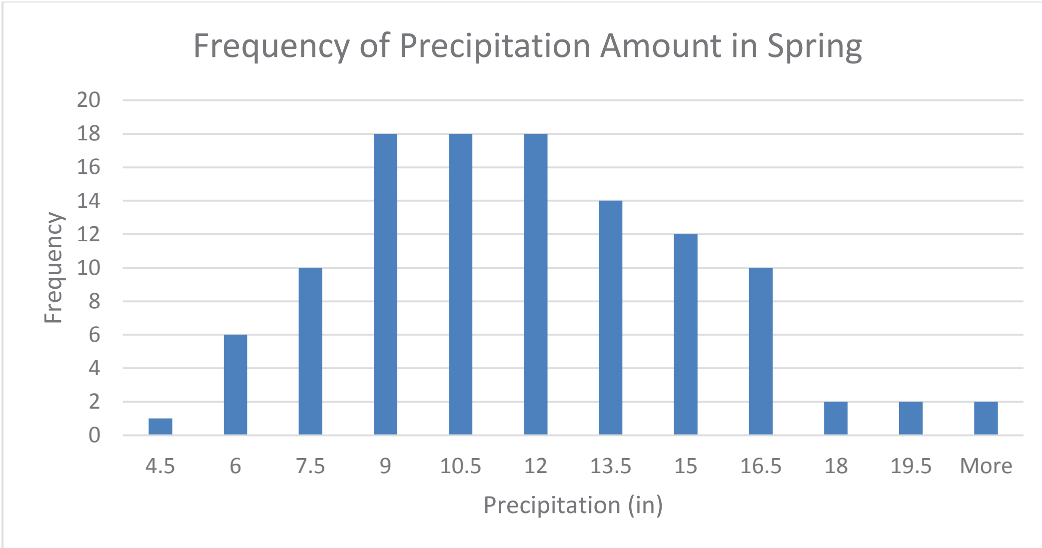

Figure 3. Histogram of Frequency of Precipitation Amount in Spring

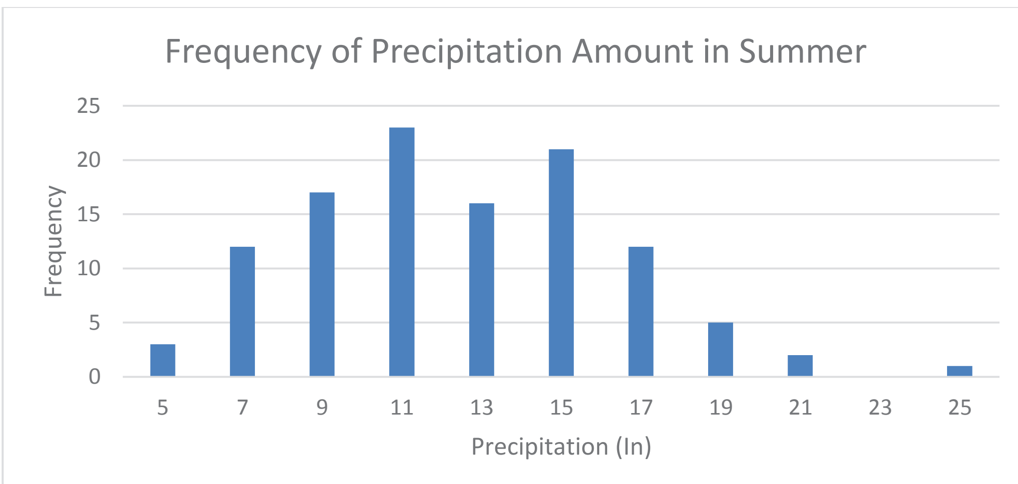

Figure 4. Histogram of Frequency of Precipitation Amount in Summer

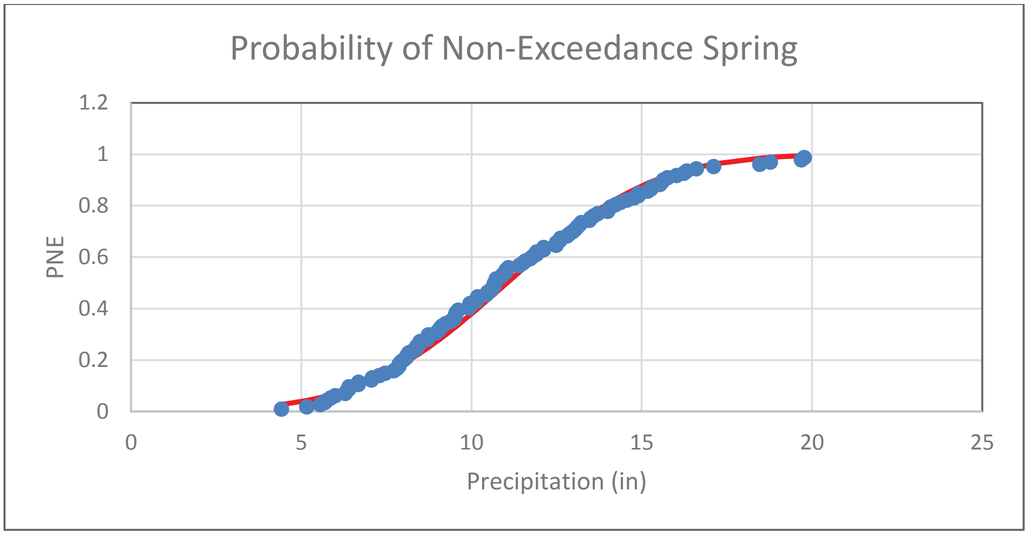

Figure 4. Frequency of Precipitation Amount in Spring

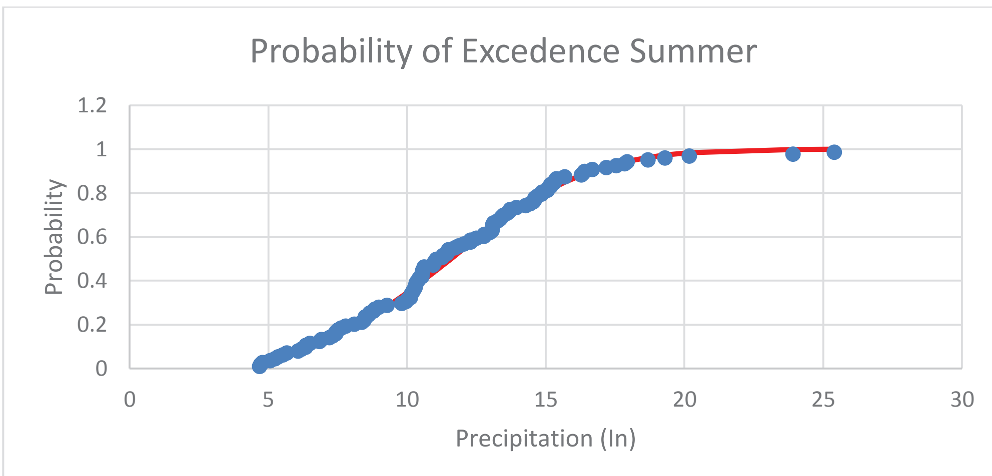

Figure 4. Frequency of Precipitation Amount in Summer

Discussion and Conclusions

In conclusion, there are many significant trends in spring and summer rainfall in Urbana. In regards to changes in the mean or variance in precipitation, there is no significant change in the spring data, but there is a significant change in the summer data. When the P two-tail value is greater than .05, we accept the null hypothesis, meaning there is no significant difference. However, when the P two-tail value is less than .05, we reject the null hypothesis, meaning that there is a significant difference. The t-Test for spring resulted in the P two-tail value of about 0.20 and the t-Test for summer resulted in the p-value of 0.006, which explains why there was a significant change in summer but not in spring. When computing the t-Test, an extra data cell was deleted from the top because from the data of the differences column, the data used for the test can't start from the beginning of the column, even if it was the greatest difference in the column. Deleting that particular data cell and finding the next highest value would determine the correct value for the P two-tail by finding the next highest amount and conducting the t-Test.

In regards to trends in precipitation, after the data has been graphed, it was determined that there is no trend in the amount of precipitation throughout the years. The R^2 values for both spring and summer are very close to zero, therefore, there is no correlation between the precipitation amount and the year. In the spring, the precipitation amounts of 9, 10.5, and 12 inches occurred the most frequently at 18. In the summer, the precipitation amounts of 11 and 15 inches occurred the most frequently at 24 and 21 respectively. When graphing the histogram, the values of precipitation amounts were rounded respectively so that it followed the pattern of increasing 1.5 inches. This would either increase or decrease the actual frequency at the correlating amounts of precipitation, depending whether it was rounded up or down to fit the pattern.

In regards to changes in the mean or variance of peak rainfall, the P two-tail value for spring was 0.4 and for summer, it was 3.6E-07. With these values, it is determined that there is no significant difference in the data for spring and there is a significant difference in the data for summer.

Finally, in regards to the changes in the number of rain days, the P two-tail value for spring was 0.0004 and for summer, it was 7.25E-09. It is then determined that both spring and summer have a significant difference in their data.

After the tests for spring and summer rainfall, the probability of non-exceedance is determined. After both the non-exceedance probability of each data value and cumulative distribution function for the fitted logNormal Distribution have been plotted, the result was almost identical. The graphs of the probability of non-exceedance and the fitted logNormal Distribution overlapped, meaning that the changes that any particular average precipitation amount will be exceeded are roughly the same as the chances of the 1901-2014-reference period of Urbana's rainfall data.

References

University of Illinois Board of Trustees. 2015. Illinois State Climatologist Data Available at: http://www.isws.illinois.edu/data/climatedb/. Accessed 27 January 2015.