MST364MidetermProject

Tianen (Benjamin) Liu

3/18/2019

1. Introduction and Motivation

All river heights change every moment due to various reasons, including temperature, strength of wind, and rainfall. Since these factors vary year by year, river height reaches different maximum values in different years. The height of the Cape Fear River near Fayetteville, NC is crucial to the agriculture, wildlife, and the homes along it. While small changes in river height may not draw anybody's attention, changes in the maximum river height in each year could cause catastrophic damage. After all, it is the highest river levels that result in serious floods. The data on the maximum river height and the proportion of the days that the river height exceeds 25 feet are of interests in various aspects. The agriculture industry and wildlife protection organizations need these data to make decisions in response to the changes that might cause damage on crops or habitat of wild animals. Local government needs these data to forecast any potential floods. Insurance companies need these data to provide the best plan for buyers as well as to maximize the company's benefits. Given an accurate model of daily log river height and a dataset of river heights in year 2017, this report will answer questions pertaining the likelihood of certain observations of river height and the whether there is evidence to believe certain changes of river heights happened in 2017 compared to historical values.

2. Methods

This section of the report describes the methods used to answer the questions mentioned at the end of Section 1. Section 2.1 is about similating river heights to find whether an observation of yearly maximum river height and a proportion of days when river height exceeds 25 ft are likely to happen or not. Section 2.2 is about using hypothesis tests to find whether the 2017 observations of the maximum river height and the proportion of days when river height exceeds 25 ft increased or not compared to historical values.

2.1 Simulating River Heights

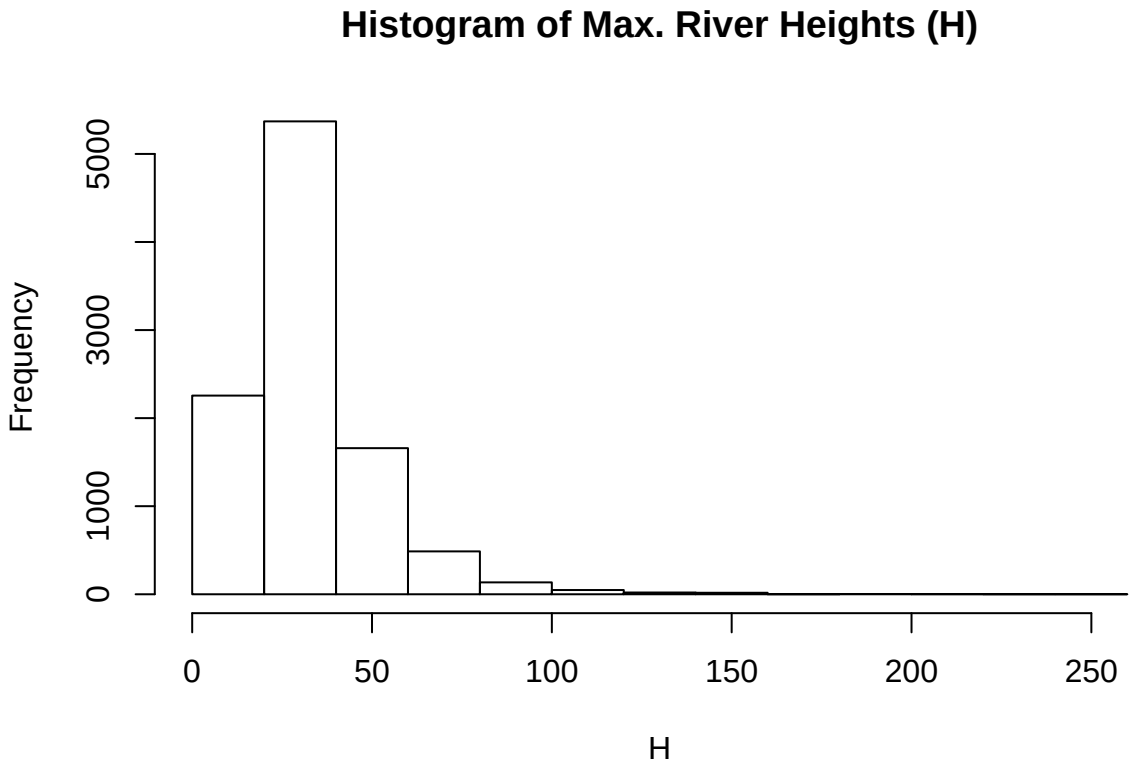
Is it unusual to observe A maximum yearly river height of at least 90 ft? Is it unusual to have A year where 20% of the days when river height exceeds 25 ft? These 2 questions are of our interest because a very high river height and a great proportion of days when river height exceeds 25 ft could mean a higher likelihood of flood.

To find out, we simulate 10,000 observations of maximum yearly river heights and 10,000 observations of proportions of days when river height exceeds 25 ft. To begin with, we assume at the start of the year, the log of river height $X_1 = log(11.5)$, and we are given an accurate model of daily log river height:

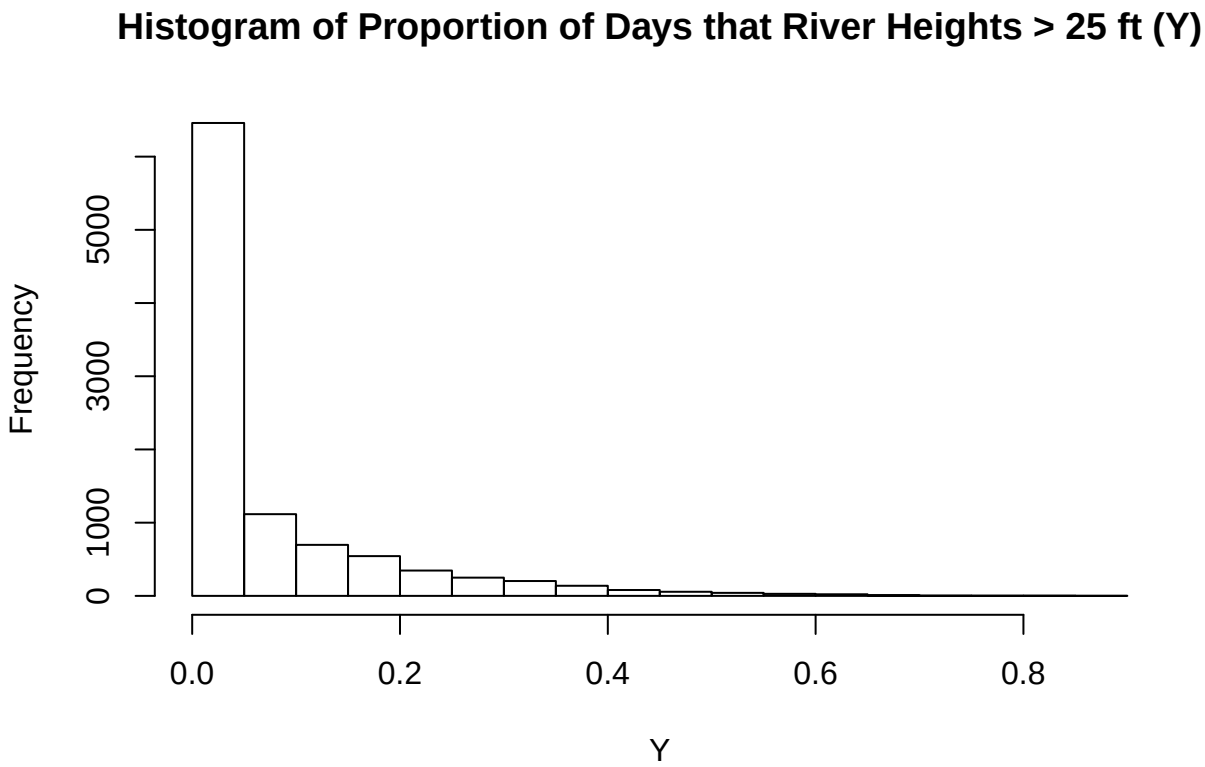
$$X_t \sim N(log(9) + 0.99(X_{t-1} - log(9)), \sigma^2 = 0.1^2)$$

Then we simulate $X_2, ..., X_{365}$ and we get the yearly river height data. Find maximum among e^{X_i} for $i = 1, ..., 365$ to get the maximum river height of this year. Call this `H`. Find the ones that are greater than 25 among e^{X_i} for $i = 1, ..., 365$ and then we can easily calculate the proportion of river heights that exceed 25 ft in this year. Call this proportion `Y`. Repeat 10,000 times to get 10,000 `H`'s and `Y`'s and we have the following 2 histograms showing the frequencies of each value of `H` and `Y`.

Histogram of H is as below:



Histogram of Y is as below:



With these 2 histograms, we can construct and interval [a, b] such that the probability that `H` or `Y` fall between a and b is 95% so that we can determine if an instance of `H` or `Y` is likely or not. This will be shown in detail in Section 3: Results and Conclusions.

2.2 Year 2017 Compared with Historical Values

Compared to historical values, does the data of 2017 provide evidence that maximum yearly river height (H) has increased from historical values? Does the data of 2017 provide evidence that the proportion of days where the river height exceeds 25 ft has increased? The former one is essentially asking if the maximum yearly river height in 2017 follows the distribution of the maximum yearly river height in historical data as if the model for daily log river height is correct; the latter one is essentially asking if the proportion of days where the river height exceeds 25 ft follows the distribution of this proportion in historical data as if the model for daily log river height is correct.

Since H and Y are not normally distributed, and we do not know the population mean and population standard deviation, it is not allowed to use the classical z-test.

Thus we use the hypothesis test. To compare H to its historical values, the null hypothesis is that the maximum yearly river height in 2017 follows the distribution of the maximum yearly river height in historical data as if the model for daily log river height is correct. The alternative hypothesis is that the null hypothesis is false. A large p-value implies that we fail to reject null and there is no evidence that the maximum yearly river height in 2017 increases compared to historical values.

To compare Y to its historical values, the null hypothesis is that the proportion of days where the river height exceeds 25 ft in 2017 follows the distribution of the proportion of days where the river height exceeds 25 ft in historical data as if the model for daily log river height is correct. The alternative hypothesis is that the null hypothesis is false. A large p-value implies that we fail to reject null and there is no evidence that the proportion of days where the river height exceeds 25 ft in 2017 increases compared to historical values.

3. Results and Conclusions

This section gives the actual computation using the methods in Section 2. Section 3.1 gives the answer to the question "Is it unusual to observe a maximum yearly river height of at least 90 ft?". Section 3.2 gives the answer to the question "Is it unusual to observe a year where 20% of the days when river height exceeded 25 ft?" Section 3.3 gives the answer to the question "Does the data in 2017 provide evidence that maximum yearly river height has increased from historical values?" Section 3.4 gives the answer to the question "Is there evidence that the proportion of days where the river height exceeds 25 ft has increased?"

3.1 A Maximum Yearly River Height of at Least 90 Feet

Based on the `Histogram of Max. River Heights (H)`, the histogram has 2 tails. Thus we find the 2.5% quantile and 97.5% quantile to get the middle 95%. The result shows that the 2.5% quantile and 97.5% quantile are:

```
##     2.5%    97.5%
## 12.81262 78.34213
```

90 ft is out of this range, which means the observation of a maximum yearly river height of 90 ft is unusual.

Alternatively, we can calculate the probability of observing a maximum yearly river height as extreme as or more extreme than 90, which is:

```
## [1] 0.0145
```

smaller than the 0.05 threshold, thus confirming our result that this observation is unusual.

3.2 A Year Where 20% of the Days When River Height Exceeded 25 Feet

Based on the `Histogram of Proportion of Days that River Heights Exceeds 25 ft (Y)`, the histogram has 1 tail. Thus we find the 95% quantile. The result shows that the 95% quantile is:

```
##        0%       95%
## 0.0000000 0.3178082
```

0.2 (= 20%) is inside this range, which means the observation of a proportion of days where the river height exceeds 25 ft is not unusual.

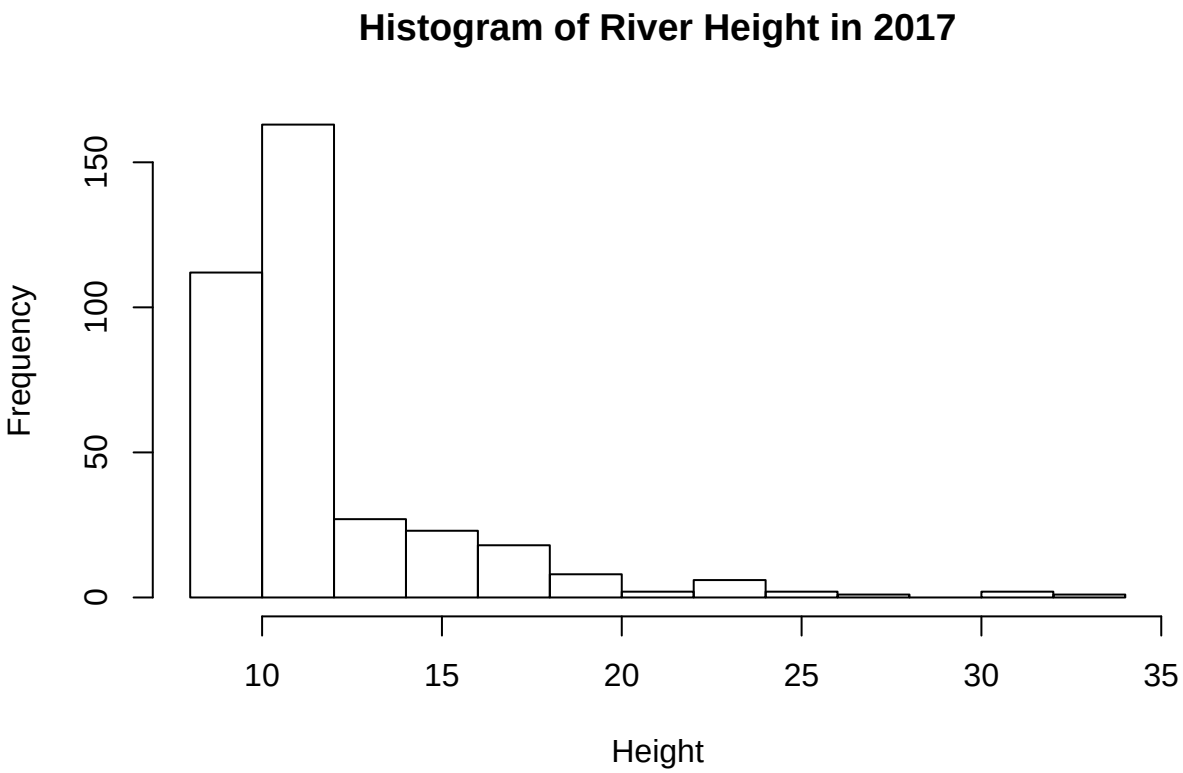
Alternatively, we can calculate the probability of observing a proportion of days where the river height exceeds 25 ft as extreme or more extreme than 20%, which is:

```
## [1] 0.1184
```

larger than the 0.05 threshold, thus confirming our result that this observation is not unusual.

3.3 Value of H in Year 2017 Compared with Historical Values

Below is the histogram of the Height data provided in 2017.



Having all river height data in 2017, we get the maximum yearly river height in 2017 as below:

```
## [1] 32.26
```

Use hypothesis test described in Section 2.2 to decide whether the maximum river height increases or not compared to historical data. Then we calculate the p-value, i.e., the probability of observing a maximum height as large as or greater than 32.26 in our simulation. The p-value is as below:

```
## [1] 0.3905
```

p-value is larger than the 0.05 threshold, which means we fail to reject the null hypothesis and conclude that there is no evidence that the maximum yearly river height in 2017 increases compared to historical values.

3.4 Value of Y in Year 2017 Compared with Historical Values

Having all river height data in 2017, we get the proportion of days where the river height exceeds 25 ft in 2017 as below:

```
## [1] 0.01369863
```

Use hypothesis test described in Section 2.2 to decide whether the proportion of days where the river height exceeds 25 ft increases compared to historical data. Then we calculate the p-value, i.e., the probability of observing a portion of days where the river height exceeds 25 ft as large as or greater than 0.01369863 in our simulation. The p-value is as below:

```
## [1] 0.5037
```

p-value is larger than the 0.05 threshold, which means we fail to reject the null hypothesis and conclude that there is no evidence that the proportion of days where the river height exceeds 25 ft in 2017 increases compared to historical values.

4. Evaluating Questions of Interest

For our questions of interest, we have the following conclusion. First, it is unusual to observe a maximum river height of at least 90 ft. Therefore, any worry about a severely high river level as 90 ft is not necessary. Second, it is not unusual to have a year where in 20% of the days river height exceeds 25 ft. Therefore, more protections should be made for places where a 25-feet river level could cause damage; and insurance companies should adjust plans for these areas accordingly. Third, there is no evidence that the maximum yearly river height in 2017 increases compared to historical values. Fourth, there is no evidence that the proportion of days where the river height exceeds 25 ft in 2017 increases compared to historical values. Therefore, industries along the Cape Fear river do not have to worry about serious change in river activities compared to previous years.