

Statistics of One Variable Assignment

Jin Hyung Park

MDM4U

Virtual High School

2020.08.07

Question 1.

Raw data is collected from Rockall- the wind speeds, measured in km/h on the fifty days in the study are recorded, in ascending order, in the table below:

a) Find the mean, mode and median wind speeds, and comment on your findings.

- To begin with, the data is used to create a frequency table.

Rockall Wind Speed Data (km/h)	
Wind Speed (km/h)	Frequency
23	1
24	1
26	1
29	1
30	2
34	1
35	1
38	1
39	1
42	1
46	1
47	1
48	1
49	2
50	1
51	2
53	2
55	1
56	1

58	1
59	1
61	1
62	1
64	1
65	1
68	3
69	1
70	1
73	1
74	2
75	1
77	1
81	2
82	1
85	1
86	1
89	1
90	2
91	1
94	1
95	1

- Calculate the Mean.

- $Mean : \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$
- $\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{41}$
- $\bar{x} = 60.18$

- The mean wind speed in Rockall on the fifty days according to the study is approximately 60 km/h.
- Calculating the Mode.
 - The mode is a value which appears most often on the given data set.
 - The mode wind speed of Rockall on the fifty days according to the study is 68km/h, which was recorded three times.
- Calculating the Median.
 - The median is a value which lies at the middle number of the given data set.
 - If a survey has an even number of datum, we can calculate the median by getting the average of two middle numbers - 25th and 26th numbers.
 - 25th number = 59
 - 26th number = 61
 - $Median = \frac{59+61}{2}$
 - Median = 60
 - The median wind speed in Rockall on the fifty days according to the study is 60km/h.
- Comment
 - Since the mean, median, mode does not have any significant differences in value, we can appreciate the fact that the distribution of wind speed of Rockall according to the study is close to the trend of normal distribution.

b) Use Excel to find the standard deviation of the wind speeds, and explain its relevance.

Raw Data(xi)	Mean(u)	Raw Value - Mean (xi-u)	Results to be squared (xi-u)^2	Sum of column D	Variance {(xi-u)^2}/50	Standard Deviation sqrt({(xi-u)^2}/50)
23	60.14	-37.14	1379.3796	20432.02	408.6404	20.21485592
24	60.14	-36.14	1306.0996			
26	60.14	-34.14	1165.5396			
29	60.14	-31.14	969.6996			
30	60.14	-30.14	908.4196			
30	60.14	-30.14	908.4196			
34	60.14	-26.14	683.2996			
35	60.14	-25.14	632.0196			
38	60.14	-22.14	490.1796			
39	60.14	-21.14	446.8996			
42	60.14	-18.14	329.0596			
46	60.14	-14.14	199.9396			
47	60.14	-13.14	172.6596			
48	60.14	-12.14	147.3796			
49	60.14	-11.14	124.0996			
49	60.14	-11.14	124.0996			
50	60.14	-10.14	102.8196			
51	60.14	-9.14	83.5396			
51	60.14	-9.14	83.5396			
53	60.14	-7.14	50.9796			
53	60.14	-7.14	50.9796			
55	60.14	-5.14	26.4196			

Note: The picture above only depicted the range from first row to 23th row.

According to the calculation executed by Excel, the standard deviation of the wind speed is about 20.21.

To begin with, the relevance of the standard deviation of a data set is that it can be used to determine the distribution of data and the appropriate proportion. The standard deviation of a data set that follows a normal distribution can be used to determine the number of standard deviations(z-score), with a raw value greater than or less than the average of the entire data set.

In a complete normal distribution, 68% of the data should be within that range, within the range of values set by 1 standard deviation above the mean and 1 standard deviation above it. When it comes to the aforementioned study, this is a wind speed between 40 and 80. It fits 58% of the raw value (29/50) for this range. Therefore, we can conclude that the Rockall's data set follows a fairly normal distribution trend.

c) Find the upper and lower quartiles of the data, and use them to draw a box and whisker diagram of the wind speeds. Use a scale that goes up to 130 km/h so that you can fit the second data set on it also.

- Lower Quartile

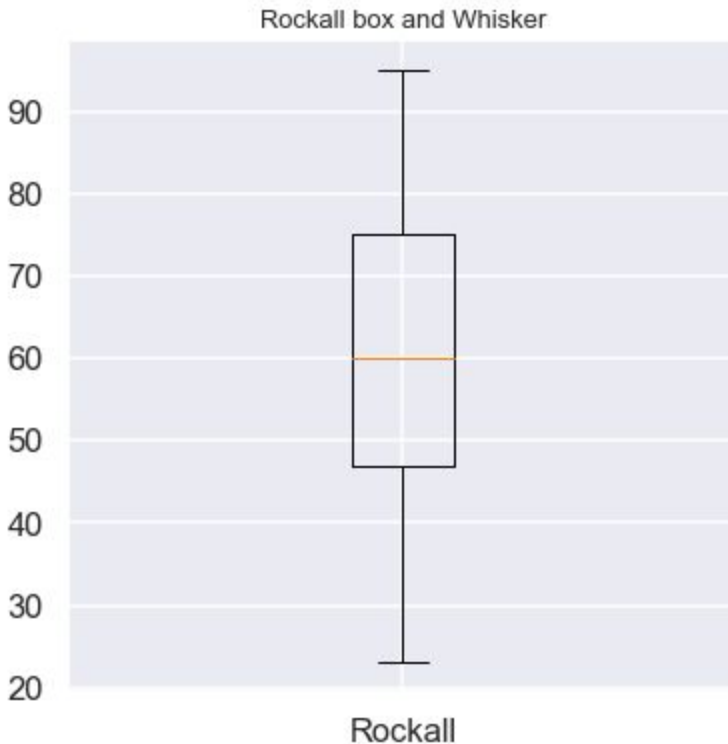
- $50 * (1/4) = 12.5$

- The lower quartile of the data is within the 13th (*rounded up*) data point which is 47.

- Upper Quartile

- $50 * (3/4) = 37.5$

- The upper quartile of the data is within the 38th (*rounded up*) data point which is 75.



Question 2.

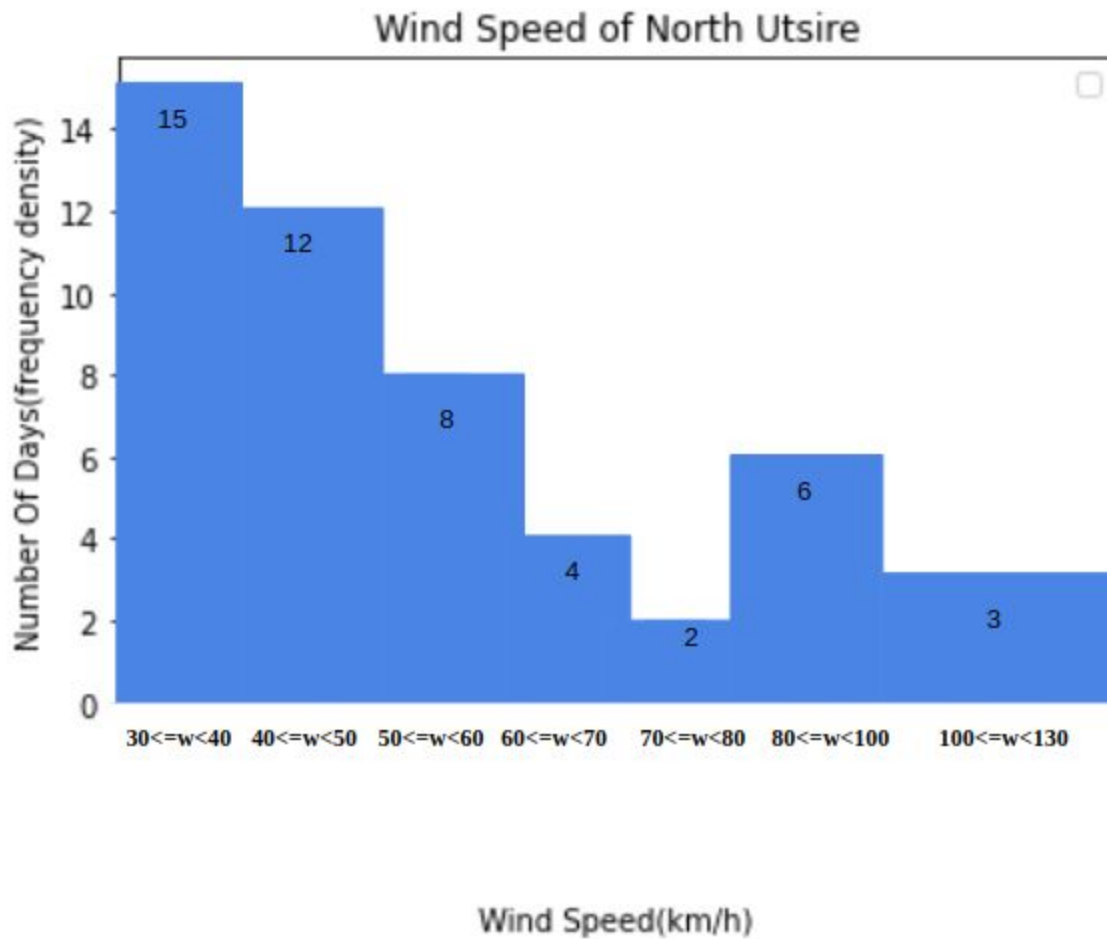
The wind speed data from North Utsire is given on the right as grouped data.

a) Find the mean wind speed.

North Utsire Wind Speed Data(km/h)					
Wind Speed (w km/h)	Frequency	Midpoint of Range	Midpoint of Range * Frequency	Sum of Results	Sum of Results divided by Sum of Frequencies
30<=w<40	15	35	525	2800	56
40<=w<50	12	45	540		
50<=w<60	8	55	440		
60<=w<70	4	65	260		
70<=w<80	2	75	150		
80<=w<100	6	90	540		
100<=w<130	3	115	345		
Sum of Frequencies	50				

The mean wind speed (km/h) in North Utsire according to the data is 56km/h.

b) Draw a histogram to represent the distribution of wind speeds.



c) Create a cumulative frequency table, and from it, draw a cumulative frequency curve. Use the curve to find the median and upper and lower quartiles.

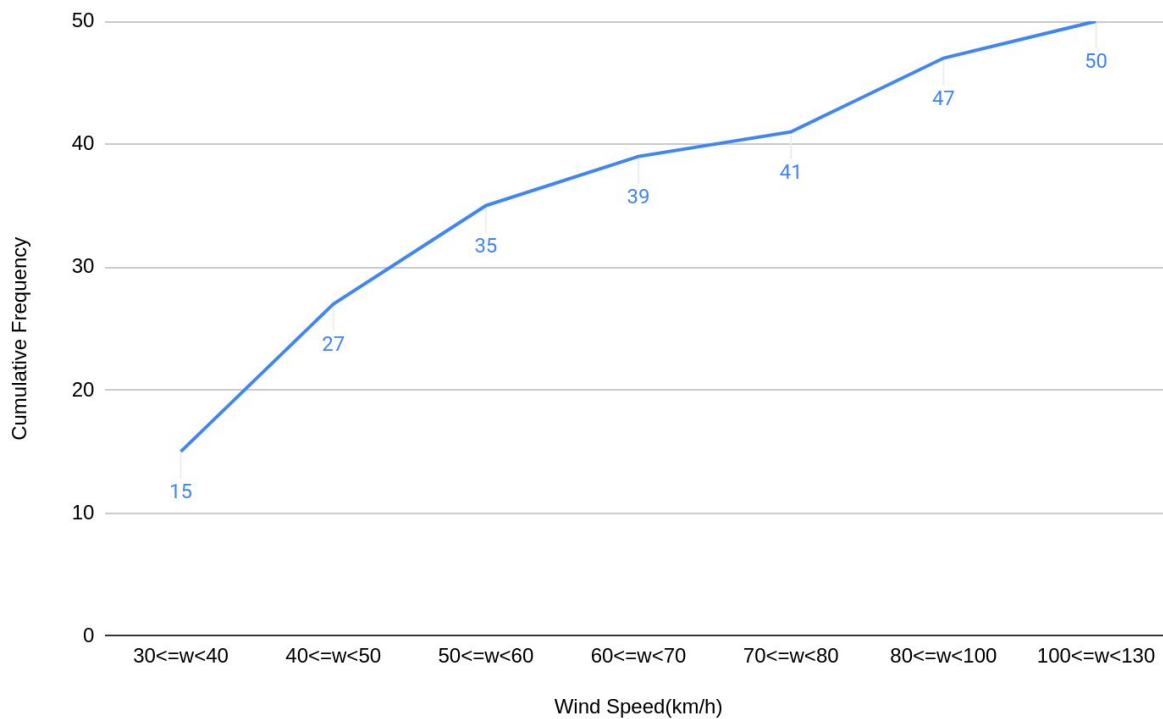
- Cumulative Frequency Table

Wind Speed (km/h)	Frequency	Cumulative Frequency
$30 \leq w < 40$	15	15
$40 \leq w < 50$	12	27
$50 \leq w < 60$	8	35
$60 \leq w < 70$	4	39
$70 \leq w < 80$	2	41
$80 \leq w < 100$	6	47

$100 \leq w < 130$	3	50
Total	50	

- Cumulative Frequency Curve:

Cumulative Frequency Curve

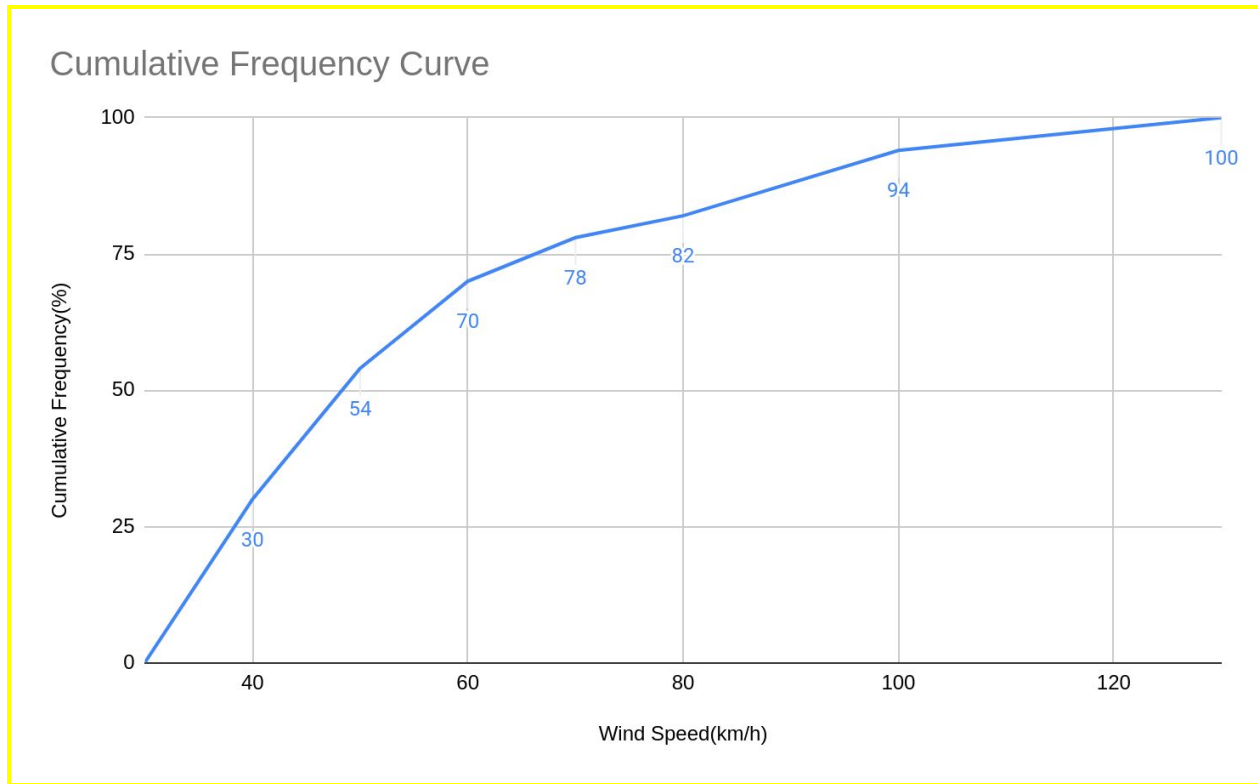


- In order to find the median and upper/lower quartiles, we need a relative frequency curve.
- Relative Cumulative Frequency Table

Wind Speed (km/h)	Relative Cumulative	Relative Frequency
30	0	0%
40	0.3	30%
50	0.54	54%
60	0.7	70%

70	0.78	78%
80	0.82	82%
100	0.94	94%
130	1	100%

- To find the median, do the following:
 - First, draw a horizontal line from the point where the midpoint of the y-axis to the point where the line intersects the curve.
 - Second, do the same thing for the lower and upper quartiles with 25% and 75% respectively.



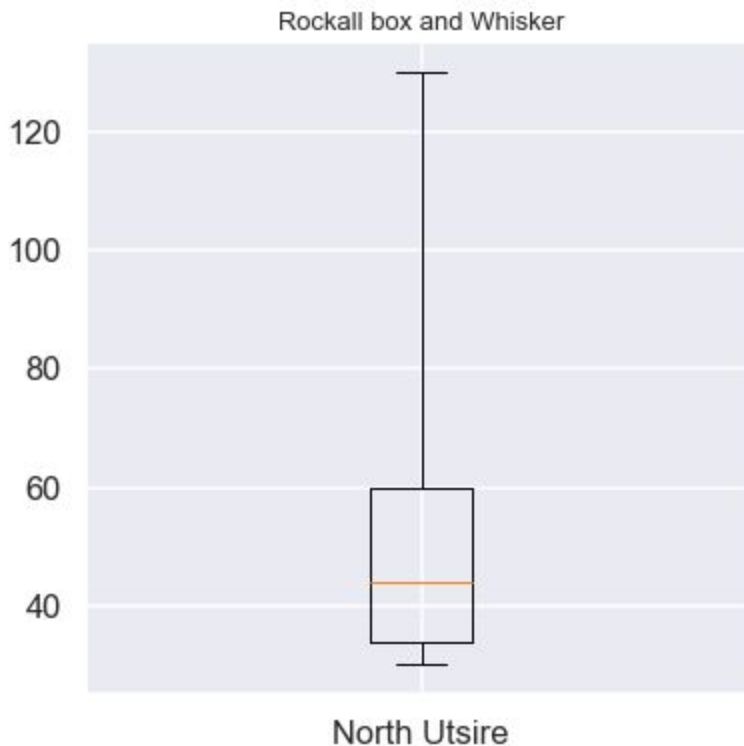
- Lower Quartile (Q1):
 - $100 \div 4 = 25\%$
 - About 39 on the x-axis goes to 25 on the y-axis.
 - The lower quartile of the data is about 39km/h.
- Median (Q2):
 - $(100 \div 2) = 50\%$
 - Two quarters of 100% is 50%.

- About 49 on the x-axis goes to 50 on the y-axis.
- The median of the data is about 49km/h.
- Upper Quartile (Q3):
 - $(100 \div 4) * 3 = 75\%$
 - Three quarters of 100% is 75%.
 - About 65 on the x-axis goes to 75 on the y-axis.
 - The upper quartile of the data is about 65km/h.

d) What do the mean, median and quartiles tell you about the distribution of wind speeds in North Utsire?

While the median of the wind speed is about 49km/h, the mean of the speed, 56km/h is much greater than the median, which is 49km/h. This clearly suggests the fact that the wind speed distribution of North Utsire is positively skewed. Since the mean is bigger than the median, the quartiles are shifted to the left which makes the distribution to be positively skewed.

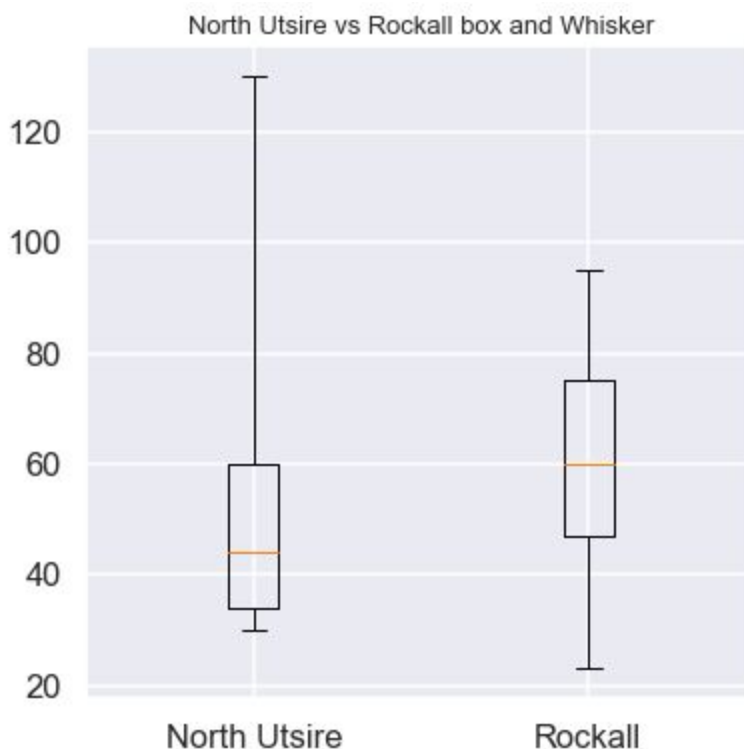
e) Use the median and quartiles to draw a box and whisker diagram on the same scale as the one for Rockall.



Question 3.

Analyze and make conclusions about Rockall and Utsire.

1. What conclusions can be drawn from the measures of central tendency for Rockall and North Utsire?
 - Rockall has an average wind speed of 60 kilometers per hour. However, the wind speed equivalent to 60 kilometers per hour is strong gust. Considering that Rockall's mode value is 68, the area will be very windy.
 - On the other hand, North Utsire is less windy than Rockall, which has an average of 56 and a median of 49. It can also be concluded that North Utsire is not as windy as Rockall because the mode interval is between 30 and 40.
2. What differences are there between the quartiles for Rockall and North Utsire? What can you conclude from these differences?
 - The Whisker box shows where 50% of the data is distributed. Local boxes are in the middle and North Utsire's boxes are skewed to the down. It can be concluded that Rockall has a normal distribution and Utsire is positively skewed.
 - Rockall's box is located between 47-75, thus the IQR (Q3-Q1) is 28, while North Utsire's box is located between 39-56, thus the IQR (Q3-Q1) is 17. This means that the mid 50% of Rockall is more widespread than that of North Utsire.
3. How is this shown by the Box and Whisker Plots? What extra information do those graphs give you?



- The box and whisker plot which is depicted above clearly shows that the distribution of Rockall is approximately located at the center of the range, while that of North Utsire has skewed right compared to that of Rockalls.
- The interquartile range means that the difference between the lower and upper quartiles. We can see that Rockall had a range of $75 - 47 = 28$ while that of North Utsire is $65 - 39 = 26$. Both have similar interquartile ranges.
- Rockall seems to have no outliers, but North Utsire seems to have upper outliers that are greater than 104km/h.

4. What overall conclusions can be made about the weather conditions in Rockall and North Utsire?

- Overall, Rockall is a consistently windy area but the wind speed is not severe. However, North Rockall is not as windy as Rockall, but it is dispersed in weather conditions which are often extreme.

5. Describe what else the various calculations and graphical representations have told you about the wind speeds in North Utsire and Rockall.

- The standard deviation of the Rockall represents the dispersion of the data set, which has an even-shaped bell curve with a standard deviation of 20.2 km/h. Under the assumption of normal distribution, it can be said that about 68% of the data will be 40 to 80 km/h.
- The cumulative frequency curve of North Utsire determines the percentile of a particular data point. For example, about wind speeds of 88 km/h will be the 90th percentile.

6. Draw comparisons between the two weather stations.

- When it comes to the case of Rockall, the candidate should include coastal areas that have huge lakes or sea. Example bounds on Five Great Lakes across Canada and the US where it has a high average of wind speed in general but does not have severe pattern changes. Gangwon Province can shed another light for the case. The region is near the East Sea of South Korea(which is also well-known as the Sea of Japan).
- For potential candidates of North Utsire, they are expected to have some cases of severe weather conditions like hurricanes or blizzards for summer and winter seasons but mild in spring and fall seasons. The candidates would be Nova Scotia or Masan/Busan province next to the Southern Sea of South Korea.