

1. Sampling method/ collection of data

To collect my data, I used the systematic sampling method. The goal for my research was to find how far people in my contacts live from their school. In order to select a sample of 16 contacts in my phone, I first made a list of all the phone numbers. There was a total of 64 contacts.

The next step was to avoid bias and ensure that every person in my contacts was identified and had an equal chance of being selected. To do this, I divided the total number of contacts I had in my phone (64) by the sample size of 16. This gave me the number 4. Which meant that I would then select every fourth person from the list. **Wow! Nice**

The final step was to choose a starting point. I used a table of random numbers and chose the first number I came across that was between 1 and 4. This gave me the number 3 which would be my starting point and I chose every fourth phone number until I had a total of 16. I called each person who was selected to ask how far they lived from their school and collected my data.

Though I did my best to give everyone in my contacts an equal chance to be selected, I was aware of some of the ways bias could be introduced into the sample. This includes the possibility that they are more likely to live near me since they know me thus leading to bias the data to a specific distance. Another problem that could occur would be that every fourth person selected lives really far from school and all the others lived a lot closer which would be unrepresentative of the population. **not necessarily**

In collecting my data, I did everything I could to avoid bias and through my data I could confirm that there was no obvious signs bias occurred.

Great!

2. Individual Data summary

The data below is listed in ascending order. **(A)** Shows the Mean of the individual data. **(B)** shows the Median. **(C)** shows the Mode which there were none. **(D)** represents how to find the standard deviation by hand. **(E)** Shows the Variance of the data and **(F)** shows the five number summary and is color coordinated to match the value on the column to the right.

0.8	0.9	1.9	2.3	2.8	3	3.3	4
6	7	7.8	9	10	11.4	12	18

Sample Size $n = 16$

A) mean: $\frac{\text{Sum of data}}{n} = \frac{0.8 + 0.9 + \dots + 18}{16} = 6.26$

B) Median: (middle data value) $\frac{4+6}{2} = 5$

C) mode: # Repeated most = none

D) Standard deviation: $S = 4.84$ $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$

X	$x - \bar{x}$	$(x - \bar{x})^2$
0.8	-5.46	29.81
0.9	-5.36	28.73
1.9	-4.36	19.00
2.3	-3.96	15.68
2.8	-3.46	11.97
3	-3.26	10.63
3.3	-2.96	8.76
4	-2.26	5.11
6	-0.26	0.07
7	.74	0.55
7.8	1.54	2.38
9	2.74	7.51
10	3.74	13.99
11.4	4.74	22.47
12	5.74	32.95
18	11.74	137.83

S = Standard deviation

$\sum$ = sum

X = Datapoint

$\bar{x}$ = mean (6.26)

n = # of datapoints (16)

$\sum = \frac{347.44}{16-1} = 4.84$

E) Variance: $S^2 = 4.84^2 = 23.4$

F) MIN = 0.8

MAX = 18

$Q_1 = \frac{2.3 + 2.8}{2} = 2.55$

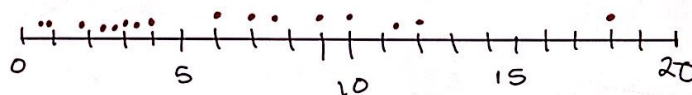
$Q_2 = \frac{4 + 6}{2} = 5$

$Q_3 = \frac{9 + 10}{2} = 9.5$

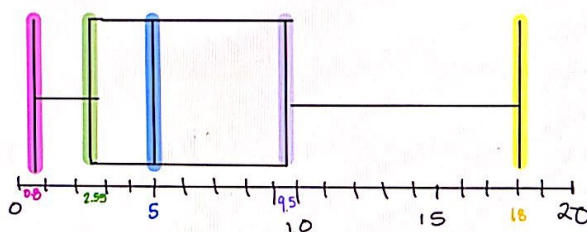
Range = MAX (18) - MIN (0.8) = 17.2

0.8
0.9
2.3
2.8
3
3.3
4
6
7
7.8
9
10
11.4
12
18

A DOT PLOT OF DATA:



A BOX PLOT OF DATA:



MIN = 0.8
MAX = 18
 $Q_1 = 2.55$
 $Q_2 = 5$
 $Q_3 = 9.5$

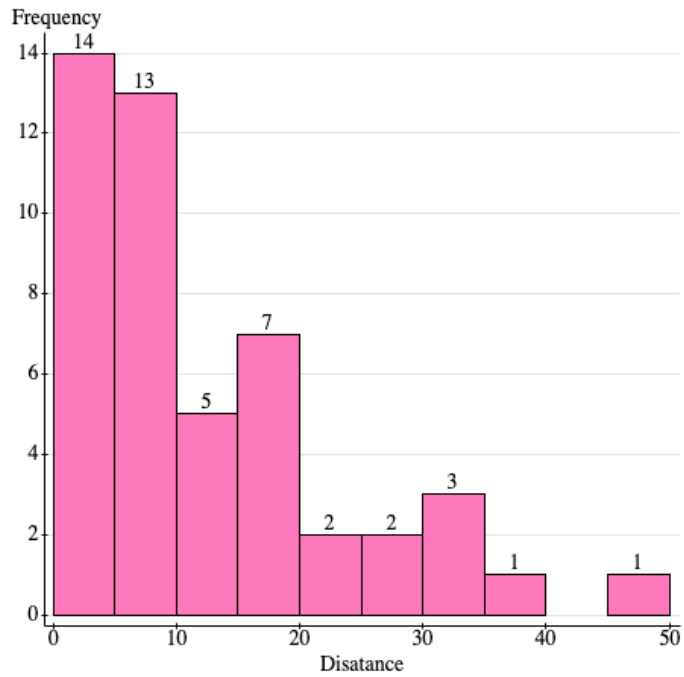
3. Group Data Summary and graphs, charts and tables

This table shows the frequency table for group 4. It includes frequency distributions, relative frequency distributions, cumulative frequency distributions and the summary statistics for all 48 data values.

Summary statistics:											
Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
var1	48	12.01875	117.40922	10.835553	1.5639774	8.65	44.8	0.2	45	3.15	17.55

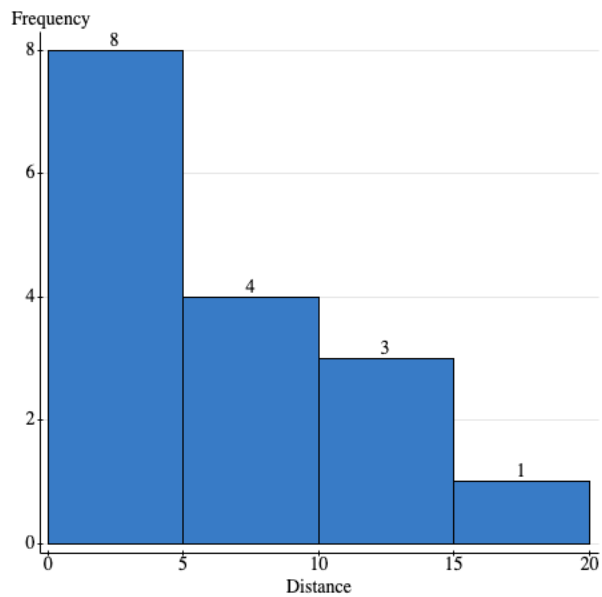
distance	Frequency	Relative Frequency	Cumulative Frequency
0-4.9	14	0.29	14
5-9.9	13	0.27	27
10-14.9	5	0.1	32
15-19.9	7	0.15	39
20-24.9	2	0.04	41
25-29.9	2	0.04	43
30-34.9	3	0.06	46
35-39.9	1	0.02	47
40-44.9	0	0	47
45-49.9	1	0.02	48

The histogram of group 4 data



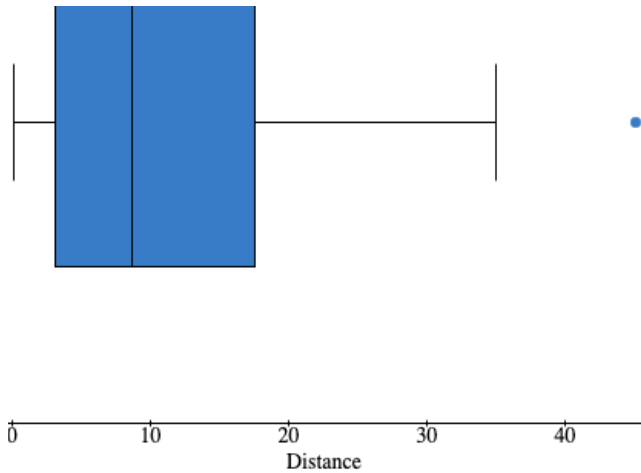
- This is a skewed histogram. It is also skewed to the right because it has a tail (shorter side) on the right and a higher side on the left.

The histogram for individual data



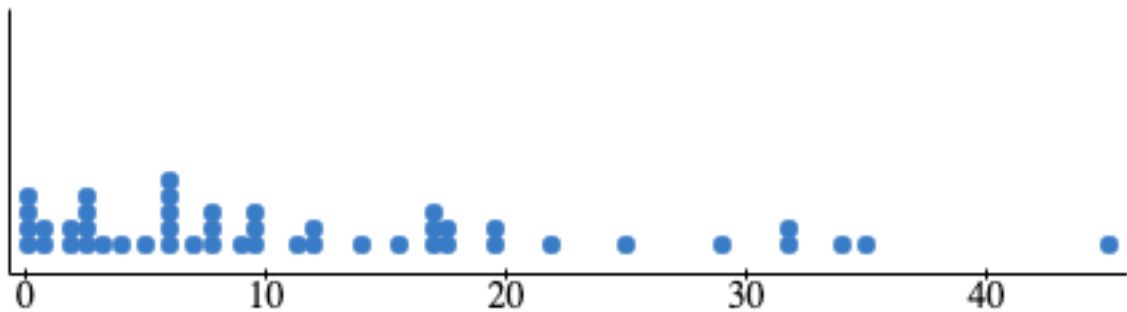
- This histogram is also skewed to the right.

Box plot for group 4



- For the box plot, the minimum distance is 0.2 and the maximum distance is 45. Q1=3.15, Q2=8.65, and Q3=17.55

Dot plot for group 4



4. Data Analysis for group 4

1. Of the 48 people, 21 live more than 10 miles from school. $21/48=0.4375$ This is 44% of the group sample.
2. Of the 48 people, 14 live less than 5 miles from school. $14/48=0.2917$ Therefore 29% of students live less than 5 miles from school.
3. $3+3.3=6.3$ $6.3/2=3.1$ Therefore 25% of students drive less than 3.1 miles This is also Q1. Also found in summary statistics.
4. $\bar{x}=12.02$ $s=10.84$
 $\bar{x}+2s=12.02+2(10.84)=33.7$ The max usual value is 33.7 miles and this concludes that my data includes three outliers. 34, 35 and 45 miles.
5. Three students out of the 48 live more than the max usual. This proportion is 6%.
6. No. It is not normal because they are not symmetrical. It is skewed to the right because the max is on the left side and the tail(shorter bars) are on the right and there is the outlier.

Great Job!