1B Descriptive Stats

Contents

1B Descriptive Statistics

[ToC]

B.1 Univariate, bivariate, and multivariate data

- Univariate
- Bivariate
- Multivariate

B.2 Speed of Light Data

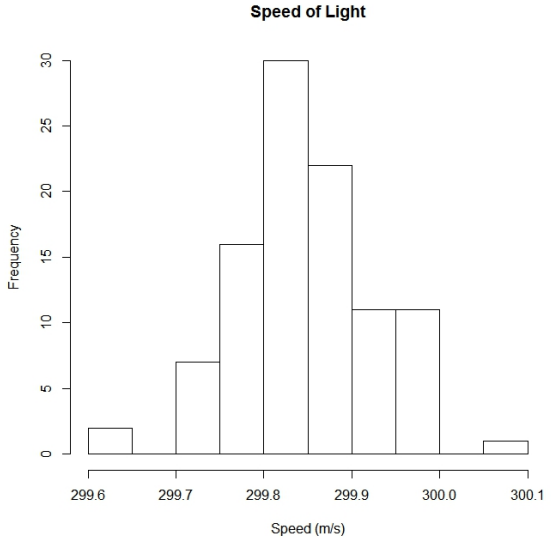
```
Light
[1] 299.85 299.74 299.90 300.07 299.93 299.85 299.95 299.98 299.98
[10] 299.88 300.00 299.98 299.93 299.65 299.76 299.81 300.00 300.00
[19] 299.96 299.96 299.96 299.94 299.96 299.94 299.88 299.80 299.85
[28] 299.88 299.90 299.84 299.83 299.79 299.81 299.88 299.88 299.83
[37] 299.80 299.79 299.76 299.80 299.88 299.88 299.88 299.86 299.72
[46] 299.72 299.62 299.86 299.97 299.95 299.88 299.91 299.85 299.87
[55] 299.84 299.84 299.85 299.84 299.84 299.84 299.89 299.81 299.81
[64] 299.82 299.80 299.77 299.76 299.74 299.75 299.76 299.91 299.92
[73] 299.89 299.86 299.88 299.72 299.84 299.85 299.85 299.78 299.89
[82] 299.84 299.78 299.81 299.76 299.81 299.79 299.81 299.82 299.85
[91] 299.87 299.87 299.81 299.74 299.81 299.94 299.95 299.80 299.81
[100] 299.87
```

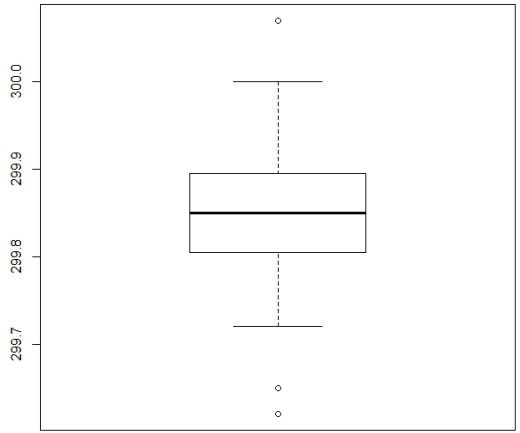
```
mean(Light)
[1] 299.8524

var(Light)
[1] 0.006242667

sd(Light)
[1] 0.07901055

summary(Light)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
299.6   299.8   299.9   299.9   299.9   300.1
```


Speed of Light
Frequency
0
5
10
15
20
25
30
299.6
299.7
299.8
299.9
300.0
300.1
Speed (m/s)


299.7
299.8
299.9
300.0

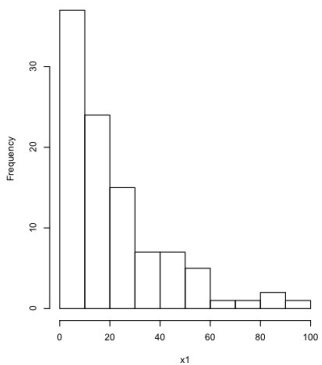
B.3 Population vs Sample

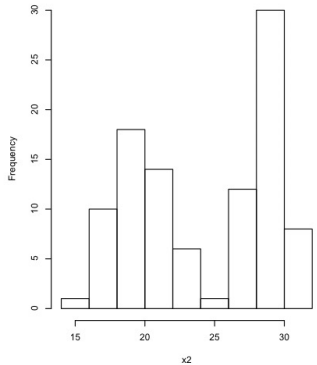
Population

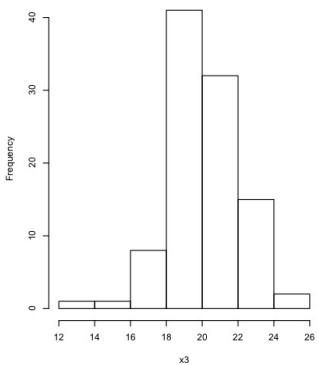
Sample

B.4 Histogram

- Unimodal, Bimodal (Multimodal), Left and right skewed.







B.5 Frequency vs Relative Frequency

Data: $1, 2, 4, 5, 3, 2, 4, 5, 1, 6$

B.6 Sample Mean, Sample Variance and Sample SD

Let $X_1, X_2, X_3, \ldots, X_n$, be a random sample of size n. Then,

Sample mean is

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Sample vaiance is

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Notice that we are dividing by $n-1$ instead of n.

Sample Standard Deviation is defined as

$$s = \sqrt{s^2}.$$

B.7 Five Number Summary of Data

is consisted of minimum observation, lower fourth, median, upper fourth, Maximum observation Boxplot

Min	Q1	Median	Q3	Max
	1st quartile	2nd quartile	3rd quartile	
	25th percentile	50th percentile	75th percentile	100th percentile

is drawn using these five numbers.

B.8 Sample median

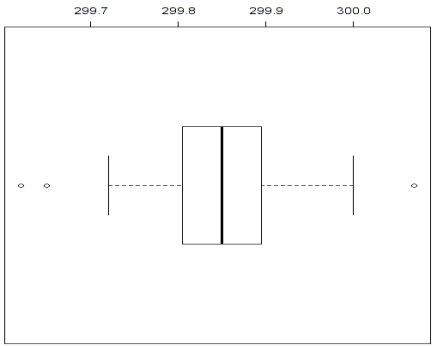
$$\widetilde{x} = \begin{cases} \frac{n+1}{2}\text{th ordred observations} & \text{if n is odd} \\ \text{average of } \left(\frac{n}{2}\right) \text{ th and } \left(\frac{n}{2}+1\right) \text{ th ordered observations} & \text{if n is even} \end{cases}$$

B.9 Get quartiles by INCLUDING median

- If Data were: $\{1, 2, 3, 4, 5, 6, 7, 8\}$

- If Data were: $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

B.10 Inter-Quartile Range is (Q3 - Q1):

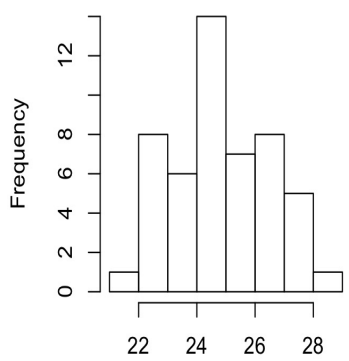


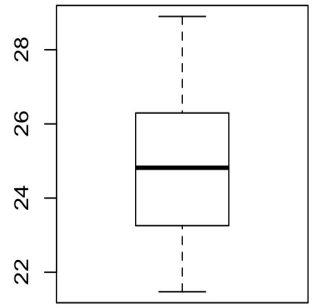
B.11 Use 5 number summary to draw a box-plot.

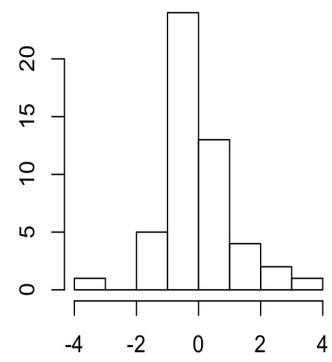
B.12 Outlier

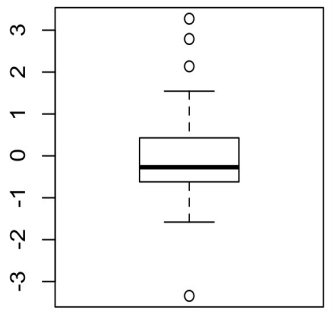
Observations farther than 1.5 box width away from the closest fourth is an outlier. If it is more than 3 box width away from the nearest fourth, it's called extreme outlier. Otherwise it is called an mild outlier.

B.13 Relate Histogram and Boxplot 1

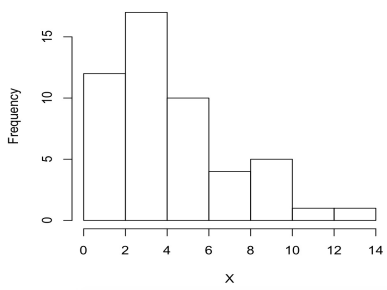

Frequency
0
2
4
6
8
12
22
24
26
28

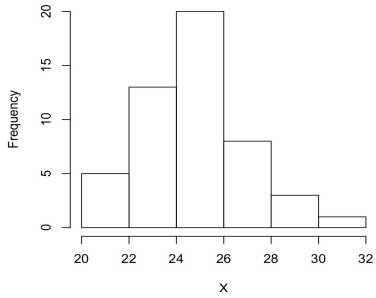

22
24
26
28


0
5
10
15
20
-4
-2
0
2
4


-3
-2
-1
0
1
2
3

B.14 Relate Histogram and Boxplot 2





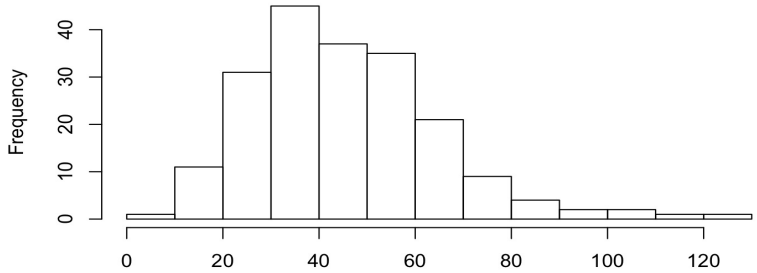
B.15 Be able to calculate variance by hand

Data : $\{2, 3, 6, 8, 9, 13\}$

i	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	2	−4.83	23.3
2	3	−3.83	14.7
3	6	−0.83	0.7
4	8	1.17	1.4
5	9	2.17	4.7
6	13	6.17	38.1
Sum	41		82.9

- $\bar{X} = 6.83$
- $Var(X) =$

B.16 Mean is larger than Median for right skewed data



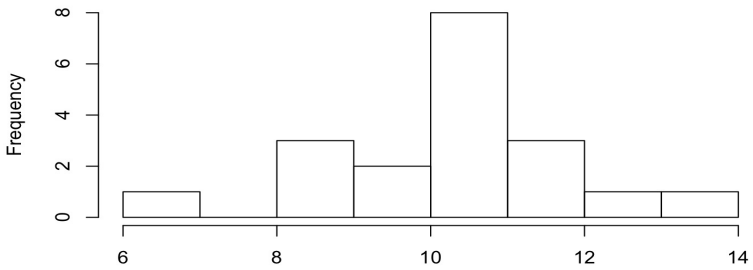
- Mean is
- Median is

B.17 Mean is much more sensitive than Median

i	X_i
1	3
2	5
3	10
4	22
5	35
Sum	75

- Mean =
- Median =

B.18 Can you get quantiles by looking at Histogram



- If $n = 19$, 1st quartile =
- Median =
- 3rd quartile =