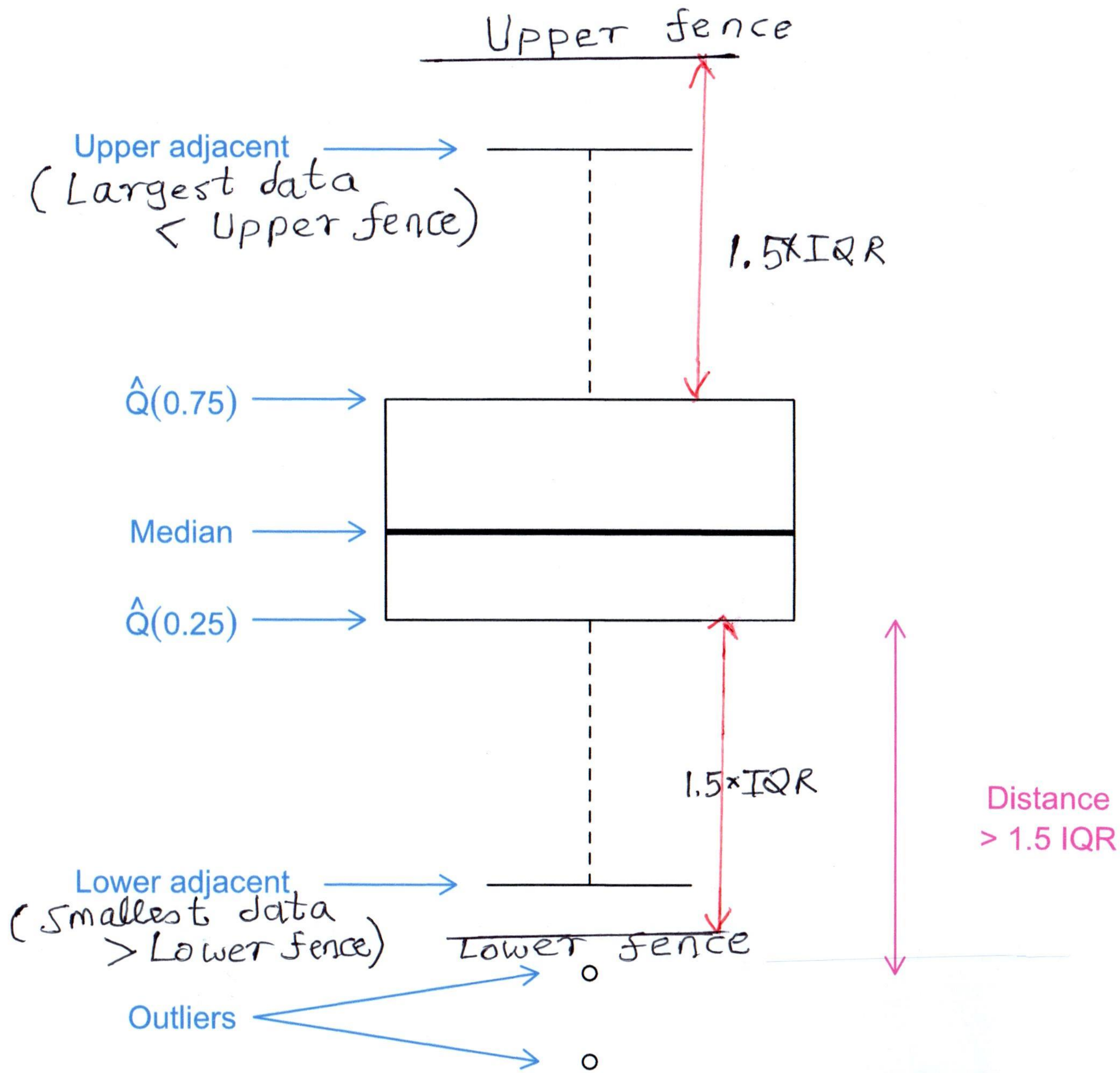


X is an <u>outlier</u> if X > upper fence
<u>OR</u> X < lower fence

R command – boxplot(X)

Example I

Data : 1, 0, 3, 2

Compute the summary statistics.

a) Order statistics

$$\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \parallel & \parallel & \parallel & \parallel \\ x_{(1)} & x_{(2)} & x_{(3)} & x_{(4)} \end{array}$$

b) $\text{mean} = \dfrac{1+0+3+2}{4} = 1.5$

c) $\text{median} = \dfrac{1+2}{2} = 1.5$

d) first quartile $Q(0.25)$

$$r' = [5 \times 0.25] = [1.25] = 1$$

$$\begin{aligned} Q(0.25) &= x_{(1)} + (1.25 - 1)\{x_{(2)} - x_{(1)}\} \\ &= 0 + 0.25 \times \{1 - 0\} \\ &= 0.25 \end{aligned}$$

e) third quartile $Q(0.75)$

$$r' = [5 \times 0.75] = [3.75] = 3$$

$$Q(0.75) = x_{(3)} + (3.75-3)\{x_{(4)} - x_{(3)}\}$$
$$= 2.75$$

f) $$s^2 = \frac{1}{4-1}\left[(0-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (3-1.5)^2\right]$$
$$= 1.666667$$ "unbiased estimate"

$$s^2 = \frac{1}{4}\left[(0-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (3-1.5)^2\right]$$
$$= 1.25$$ "biased estimate"

g) $$SD = \sqrt{1.666667}$$ "unbiased estimate"
$$= \sqrt{1.25}$$ "biased " "

h) Range $= 3 - 0 = 3$

i) $$IQR = Q(0.75) - Q(0.25)$$
$$= 2.75 - 0.25$$
$$= 2.5$$

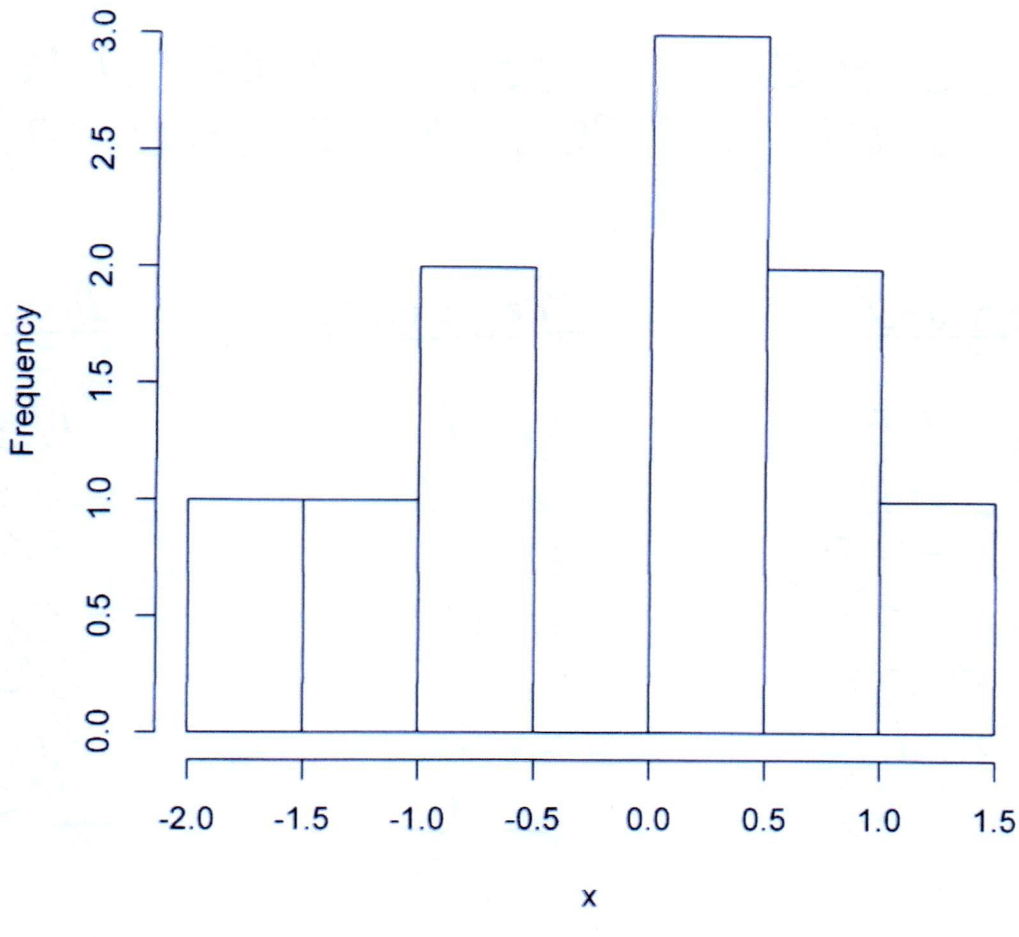
Example 2

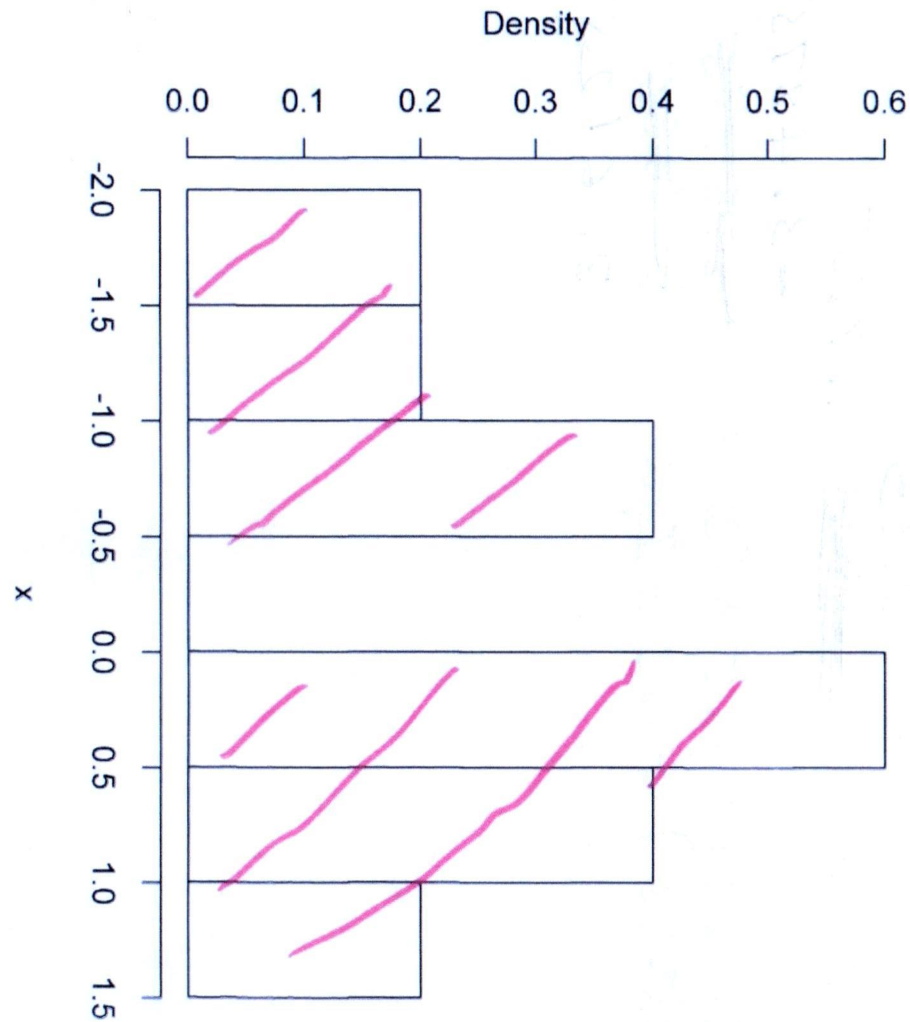
Data: 0.98, −1.20, −0.65, 0.16, 0.18
1.1, −1.8, 0.32, −0.7, 0.94

Draw the histogram with interval width = 0.5

Interval	Frequency	Density
$[-2, -1.5)$	1	1/5
$[-1.5, -1)$	1	1/5
$[-1, -0.5)$	2	2/5
$[-0.5, 0)$	0	0
$[0, 0.5)$	3	3/5
$[0.5, 1)$	2	2/5
$[1, 1.5)$	1	1/5

$$\text{Density} = \frac{\text{Frequency}}{(\text{Interval width}) \times (\text{No of data})}$$


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Frequency
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
x


Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
x

Example 3

Data : 0.98, -1.2, -0.65, 0.18, 0.16

1.1, -1.8, 0.32, -0.7, 0.94

Draw the boxplot

median = 0.17

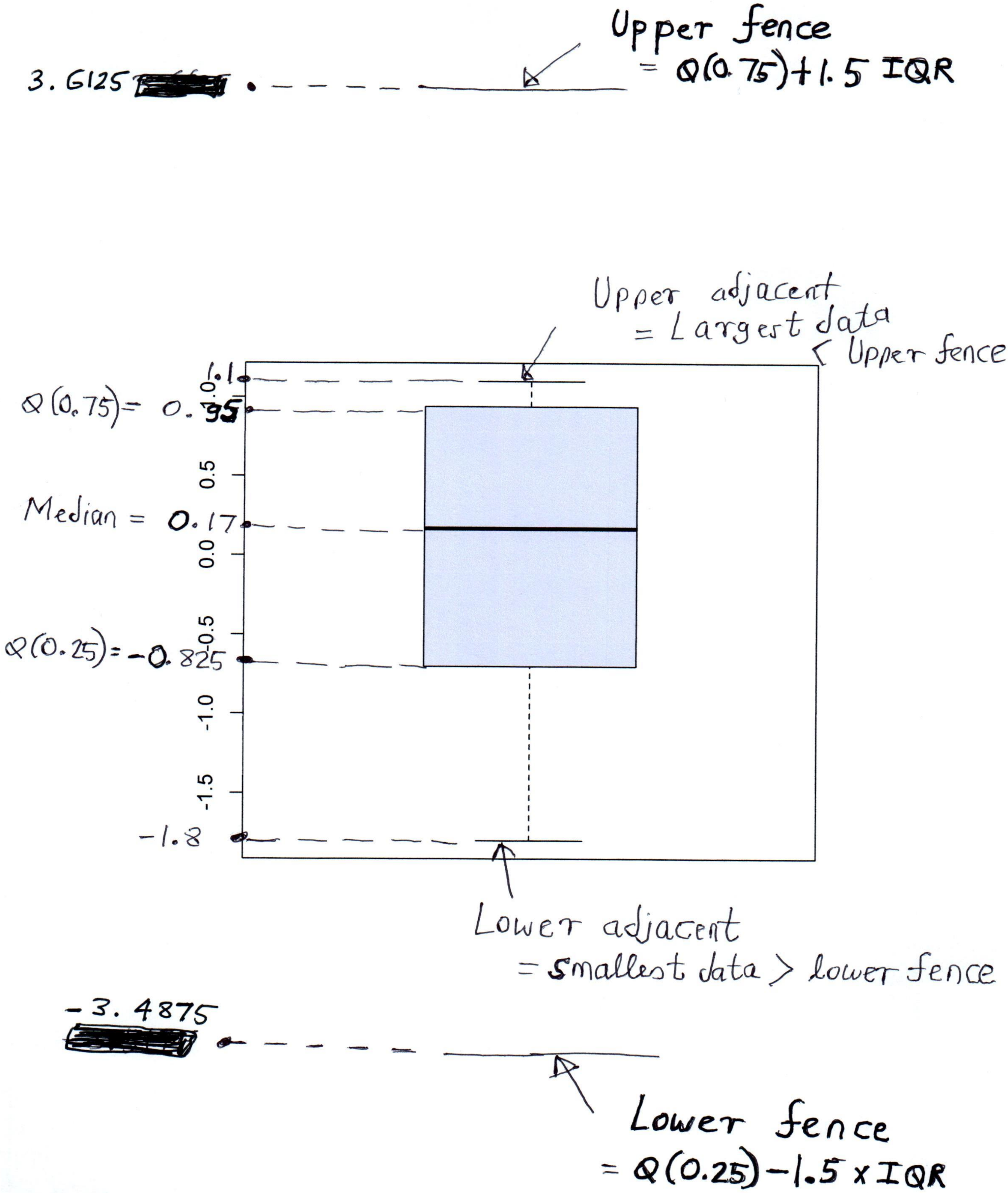
$Q(0.25)$ = - 0.825

$Q(0.75)$ = 0.95

IQR = 1.775

Lower fence = $Q(0.25) - 1.5 \times IQR = -3.4875$

Upper fence = $Q(0.75) + 1.5 \times IQR = 3.6125$


Upper fence
= Q(0.75)+1.5 IQR
3.6125 [redacted]
Upper adjacent
= Largest data
< Upper fence
1.1
Q(0.75) = 0.95
1.0
0.5
Median = 0.17
0.0
-0.5
Q(0.25) = -0.825
-1.0
-1.5
-1.8
Lower adjacent
= smallest data > lower fence
-3.4875
[redacted]
Lower fence
= Q(0.25)-1.5 x IQR

Example 4

Suppose $X_1, \ldots, X_n$ are IID with $E(X) = \mu$ and $Var(X) = \sigma^2$.

Show that

$$E\Big(\underbrace{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}_{\text{unbiased version of sample variance}} \Big) = \sigma^2$$

$\boxed{\text{2 steps}}$

(i) $\displaystyle \sum_{i=1}^{n} (X_i - \bar{X})^2$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\bar{X} - \mu)^2$$

(ii) $E\left[(X_i - \mu)^2 \right]$

$= Var(X_i)$

$= \sigma^2$

& $E\left[(\bar{X} - \mu)^2 \right]$

$= Var(\bar{X})$

$= \sigma^2 / n$