1.1.5 Frequency Count

Example 1.8. *Suppose you have a data A:* 1, 1, 1, 3, 4, 3, 5, 5, 6, 6, 6, 2, 1, 3, 7, 7, 7, 4, 3, 5, 7, 1, 2, 3, 4, 8. *Using R, how to find the frequency of a particular number? Hence draw a Frequency bar diagram (Column diagram) of the above data.*

► *R Code*:

```
a=c(1,1,1,3,4,3,5,5,6,6,6,2,1,3,7,7,7,4,3,5,7,1,2,3,4,8)
b=table(a)
b
cbind(b)
# For Column Diagram
plot(b,col='red',xlab = 'Numbers',ylab = 'Frequency',lwd=3)
```

► *R Plot*: *See the plot Fig. 1.5.*

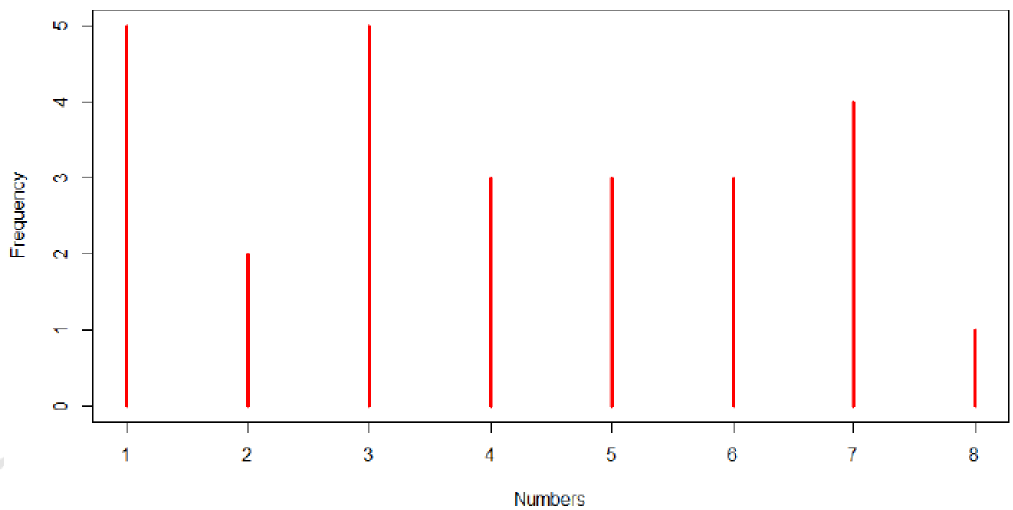

Figure 1.5: Frequency bar diagram (Column diagram) of the given data in Problem 1.8

[Do It Yourself] 1.4. *Suppose you have the data: 68 43 63 62 57 74 40 04 42 30 50 65 20 42 35 59 15 70 77 77 74 17 08 39 44 23 29 50 29 75 23 49 55 40 15 73 08 45 45 21 49 15 54 51 07 63 08 60 12 03 09 61 08 37 29 62 05 18 37 61 21 51 56 29 13 15 21 22 69 25 48 23 77 20 54 27 02 38 31 70 19 21 51 09 68 56 78 64 74 69 61 04 55 56 10 18 49 18 11 14 65 02 14 38 49 22 48 19 68 48 30 68 11 65 44 09 48 78 45 47 39 21 39 42 76 63 55 46 69 15 24 19 52 75 01 28 41 49 46 09 11 62 37 67 33 52 07 25 61 50 40 11 13 44 62 22 14 15 54 33 11 11 68 56 08 08 06 25 55 37 38 68 80 23 54 37 24 42 12 49 03 27 13 16 27 69 50 52 32 78 68 43 41 03 60 01 18 70 18 63 25 75 33 73 33 25 19 52 48 23 24 24 44 30 18 21 38 34 30 73 67 26 52 32 52 57 80 38 79 42 17 32 61 07 22 39 13 57 20 46 71 28 76 26 73 45 03 12 18 43.*
Using R, write down the frequency of a particular number. Also draw the column diagram.

1.1.6 Histogram

Example 1.9. *Suppose you have a data on heights (cm) and frequency of students in a class. Using R draw Histogram with normal approximation.*

Continuous Data	
Students Height	Frequency
144.55 - 149.55	1
149.55 - 154.55	3
154.55 - 159.55	24
159.55 - 164.55	58
164.55 - 169.55	60
169.55 - 174.55	27
174.55 - 179.55	2
179.55 - 184.55	2

Table 1.3: Height Frequency Data. [Source: Fundamental Vol. 1]

► *R Code*:

```
a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55),c(1,3,24,58,60,27,2,2))
H=hist(a,breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55))
plot(H,xlab="Height",ylab="Frequency",main="Hisogram",ylim=c(0,70))
# With normal approximation
b=length(a)
b1 = min(diff(H$breaks))
curve(b*b1*dnorm(x,mean(a),sd(a)),col="blue",add=TRUE)
```

```
# Histogram : Data by absolute numbers

a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55),c(1,3,24,58,60,27,2,2))
H=hist(a,breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55))
H$counts= H$counts/5
plot(H,xlab="Height",ylab="Frequency Density",main="Hisogram",ylim=c(0,15))
# With normal approximation
b=length(a)
curve(b*dnorm(x,mean(a),sd(a)),col="blue",add=TRUE)
```

```
# Histogram : Data by proportion

a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55),c(1,3,24,58,60,27,2,2))
H=hist(a,breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55))
H$counts= H$counts/(5*sum(H$counts))
plot(H,xlab="Height",ylab="Relative Frequency Density",main="Hisogram",ylim=c(0,0.1))
# With normal approximation
curve(dnorm(x,mean(a),sd(a)),col="blue",add=TRUE)
```

▶ R Plot: *See the plot Fig. 1.6. Here a) Height vs. Frequency (f) b) Height vs. Frequency Density (f/c) c) Height vs. Relative Frequency Density (f/nc)*

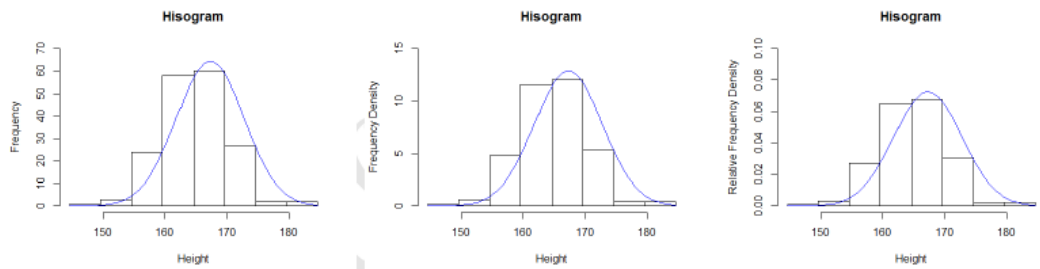

Figure 1.6: Histogram with normal approximation. Data are taken from Table 1.3. Here total area are a) nc b) n c) 1. Note: n is total frequency and c is class width.

▶ *Instead of taking upper class boundary in 'a', we can take the average class value also.*

[Do It Yourself] **1.5.** *Suppose the marks of* 100 *students in a class is as follows:*
37 52 07 52 35 52 68 42 66 82 14 59 70 75 74 38 51 38 30 69 52 64 67 53 59 61 65 68 45 41 70 25 89 42 59 52 84 45 25 64 65 71 34 77 86 81 70 74 68 72 10 15 68 30 30 57 62 34 74 51 60 27 51 77 52 54 65 27 37 68 80 94 50 36 43 44 51 54 71 35 44 55 58 47 75 28 45 46 44 45 81 12 81 92 56 60 36 65 49 78.
Draw a histogram on that data. Change the class width of your own. Also draw the normal approximation curve.

1.1.7 Box Plot

■ A box plot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points.

▶ Box plots can be drawn either horizontally or vertically.

▶ A typical boxplot looks like Fig. 1.7.

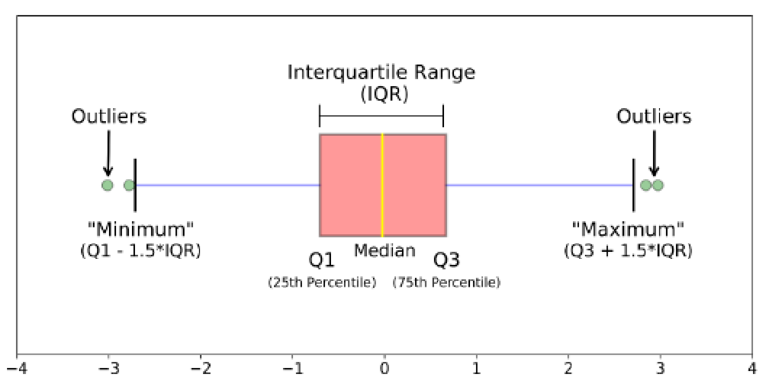

Figure 1.7: Shape of a typical Boxplot.

► A typical boxplot compared with probability distribution looks like Fig. 1.8.

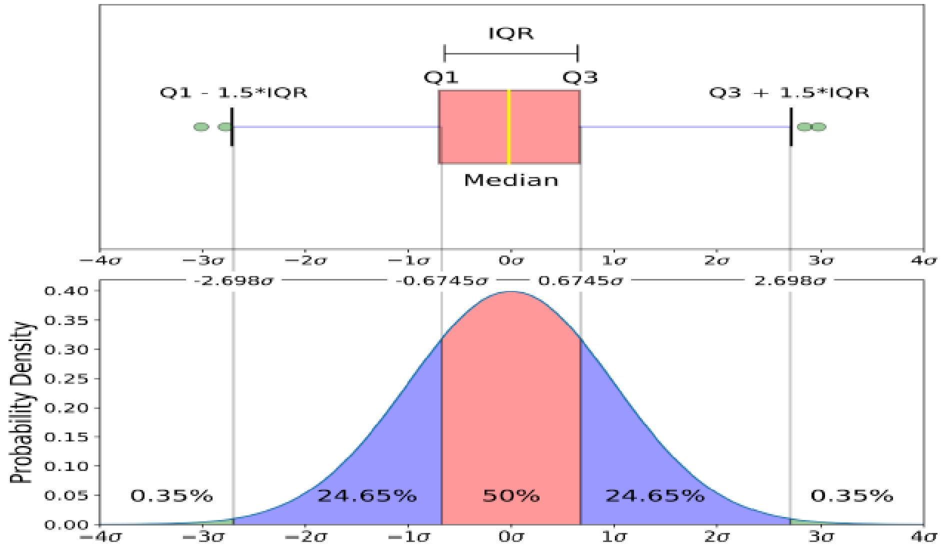

Figure 1.8: Boxplot compared with probability distribution.

Example 1.10. *Suppose you have a data on students marks in a subject. Using R draw a box plot based on the data given below.*

Discrete Data			
Student No.	Marks	Student No.	Marks
1	31	11	56
2	71	12	34
3	69	13	62
4	59	14	15
5	70	15	53
6	74	16	81
7	41	17	76
8	87	18	24
9	37	19	21
10	46	20	08

Table 1.4: Student Marks Data.

▶ R Code:

```
a=c(31,71,69,59,70,74,41,87,37,46,56,34,62,15,53,81,76,24,21,8)
summary(a)
boxplot(a)
boxplot(a,col = "red")
```

▶ R Plot: *See the plot Fig. 1.9.*

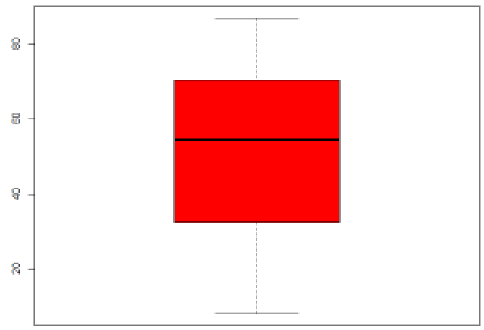

Figure 1.9: Boxplot for the data of Table 1.4.

Example 1.11. *Suppose you have a data on students marks in two subjects. Using R draw a multiple box plot based on the data given below.*

Discrete Data					
Student	Marks A	Marks B	Student	Marks A	Marks B
1	31	10	11	56	78
2	71	23	12	34	93
3	69	55	13	62	94
4	59	60	14	15	90
5	70	62	15	53	65
6	74	68	16	81	88
7	41	53	17	76	76
8	87	78	18	24	96
9	37	60	19	21	86
10	46	87	20	08	66

Table 1.5: Student Marks Data.

▶ R Code:

```
a=c(31,71,69,59,70,74,41,87,37,46,56,34,62,15,53,81,76,24,21,8)
b=c(10,23,55,60,62,68,53,78,60,87,78,93,94,90,65,88,76,96,86,66)
dat=cbind(a,b)
colnames(dat) = c("Marks A","Marks B")
boxplot(dat)
boxplot(dat,col = c("red","blue"))
```

▶ R Plot: *See the plot Fig. 1.10.*

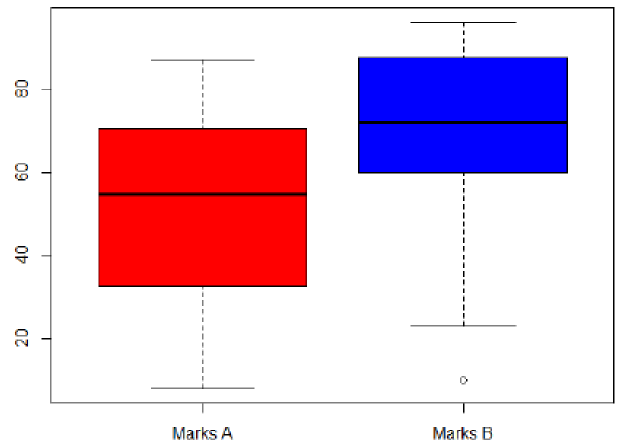

Figure 1.10: Parallel Boxplot for the data of Table 1.4.

[Do It Yourself] 1.6. *Suppose* 25 *students has marks in three different tests as follows.*
Test I: 77 69 70 65 11 06 91 88 80 81 62 83 56 75 51 45 94 99 51 87 97 97 85 82 86
Test II: 53 34 10 43 44 27 45 53 43 58 54 60 22 82 17 49 43 71 26 98 45 24 31 90 19
Test III: 56 74 72 61 64 56 75 49 55 68 72 54 56 48 63 71 71 55 68 65 02 99 56 71 42.
Using R draw a multiple box plot based on the given data.

1.1.8 Stem-leaf Plot

■ A stem-and-leaf plot of a quantitative variable is a textual graph that classifies data items according to their most significant numeric digits. In addition, we often merge each alternating row with its next row in order to simplify the graph for readability.
▶ If the data is: $a = (11, 13, 18, 16, 21, 29, 24, 30, 39, 34, 33)$, we have to divide the data into stems (left part) and leafs (right part). Hence, the plot will be as follows: 1 | 1368, 2 | 149, 3 | 0349.

Example 1.12. *Suppose you have a Data A on students marks in a subject as* 11, 13, 18, 16, 21, 29, 24, 30, 39, 34, 33. *Using R draw a stem-leaf plot based on the data. Now, for the Data B:* 11, 13, 18, 16, 21, 29, 24, 30, 39, 34, 33, 48, 47, 42, 50, 55, 53, 52, 61, 69, 65, 62 *draw stem-leaf plot and compare with the previous one.*

▶ R Code:

```
a=c(11,13,18,16,21,29,24,30,39,34,33)
stem(a)
b=c(11,13,18,16,21,29,24,30,39,34,33,48, 47,42,50,55,53,52,61,69,65,62)
stem(b)
stem(b,scale=2)
```

▶ R Plot: *The output for Data A is okay. For the output of Data B, the stems are* 0, 2, 4, 6 *i.e. they are grouped. However, after using 'scale = 2' the stems are* 1, 2, 3, 4, 5, 6 *and it is the simple form.*