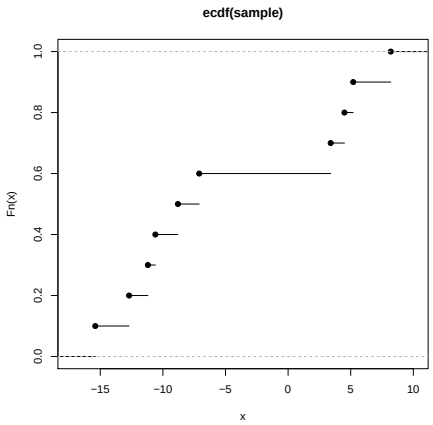
Chapter 2

Frequency distributions

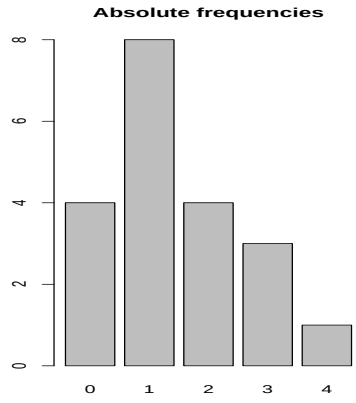
Exercise 3 *Answer* :

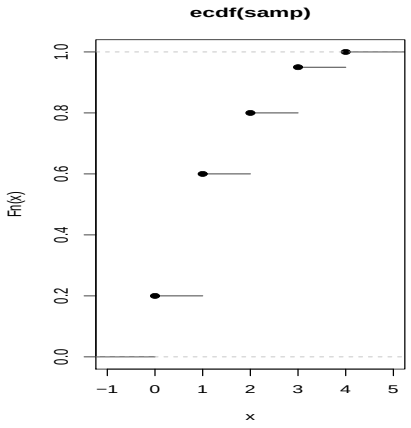
```
> sample<-c(-15.4,-8.8,8.2,3.4,-7.1,4.5,-12.7,5.2,-10.6,-11.2)
> cumsum(table(sample))/length(sample)
 -15.4  -12.7  -11.2  -10.6  -8.8  -7.1  3.4  4.5  5.2  8.2
   0.1    0.2    0.3    0.4   0.5   0.6  0.7  0.8  0.9  1.0
> plot(ecdf(sample))
```



Exercise 4 *Answer* :

```
> samp<-c(1,1,2,1,0,3,4,2,3,1,0,2,1,1,0,1,1,0,3,2)
> freq<-table(samp)
> vec_freq<-c(freq)
> data.frame(Abs.freq=vec_freq,Rel.freq=vec_freq/length(samp))
   Abs.freq  Rel.freq
0         4      0.20
1         8      0.40
2         4      0.20
3         3      0.15
4         1      0.05
> par(mfrow=c(1,2))
> barplot(table(samp),main="Absolute frequencies")
> plot(ecdf(samp))
```





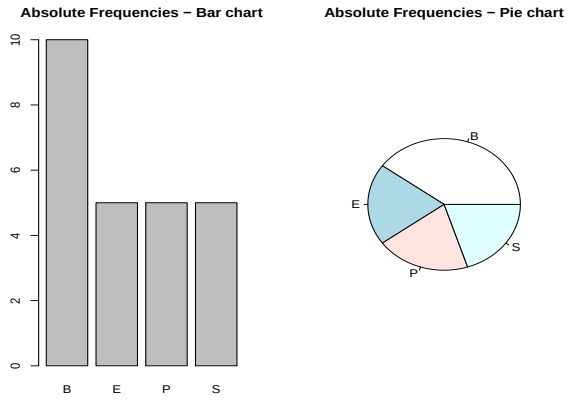
Exercise 5 *Answer* :

1. The statistical population consists in the sample of 25 students of the X-University in X-town. The units of this population in this survey are the students.
   The characteristics of identification in this population which may be defined are

   - the gender,
   - the field of study,
   - the number of siblings,
   - the income.

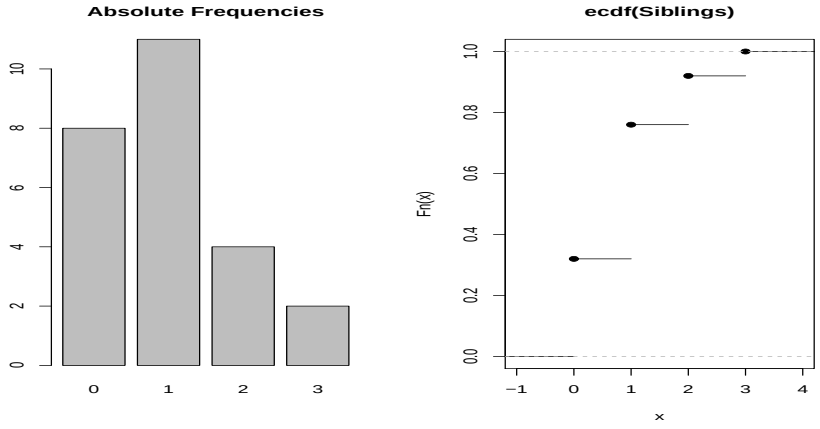
2. The category "Field of study" is nominal since it has no inherent order.

```
> Name<-c("MA","UA","WA","KB","SB","ED","KD","TE","JF","EG","KH","AK","TK","CL","UL",
+ "AM","MM","RM","BN","AR","CR","BS","CS","UT","CW")
> Studies<-c("E","S","B","B","P","P","S","E","P","B","B","E","B","S","P","B","B","S",
+ "B","B","E","B","S","E","P")
> Siblings<-c(0,1,0,1,1,2,2,1,1,0,0,1,0,3,2,0,1,0,1,2,1,1,3,0,1)
> Income<-c(924,789,1365,683,744,640,631,814,778,1062,1230,700,850,641,640,850,683,
+ 616,683,683,660,1440,794,660,640)
> Students<-data.frame(Name,Studies,Siblings,Income)
> data.frame(Abs.Freq=c(table(Studies)),Rel.Freq=c(table(Studies))/nrow(Students))
   Abs.Freq  Rel.Freq
B        10       0.4
E         5       0.2
P         5       0.2
S         5       0.2
> par(mfrow=c(1,2))
> barplot(table(Studies),main="Absolute Frequencies - Bar chart")
> pie(table(Studies),main="Absolute Frequencies - Pie chart")
```



3. The category "Number of siblings" is ratio scaled.

```
> data.frame(Abs.freq=c(table(Siblings)),Rel.Freq=c(table(Siblings))/nrow(Students))
   Abs.freq  Rel.Freq
0         8      0.32
1        11      0.44
2         4      0.16
3         2      0.08
> cumsum(table(Siblings))/nrow(Students)
    0     1     2     3
0.32  0.76  0.92  1.00
> par(mfrow=c(1,2))
> barplot(table(Siblings),main="Absolute Frequencies")
> plot(ecdf(Siblings))
```



4. In order to answer, it suffices to look at the previous table (or graph which is less accurate): $0.92 \times 25 = 23$ students have at most 2 siblings. We may also use R:

```
> ecdf(Siblings)(2)
[1] 0.92
```

5. We need to compute $1 - \tilde{F}_n(1)$:

```
> 1-ecdf(Siblings)(1)
[1] 0.24
```

24% percent of the students have at least 2 siblings.

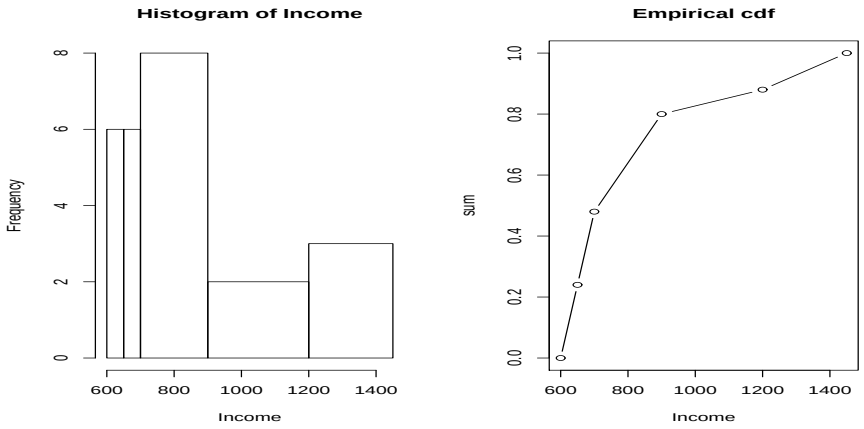
6. We need to compute $\tilde{F}_n(2) - \tilde{F}_n(0)$:

```
> ecdf(Siblings)(2)-ecdf(Siblings)(0)
[1] 0.6
```

60% percent of the students have 1 or 2 siblings.

7. The category "Income" is interval scaled.

```
> class<-c(600,650,700,900,1200,1450)
> tab<-data.frame(Abs.Freq=c(table(cut(Income,class,include.lowest=T,right=F))),
+ Rel.Freq=c(table(cut(Income,class,include.lowest=T,right=F)))/nrow(Students))
> tab
                     Abs.Freq  Rel.Freq
 [600,650)                  6      0.24
 [650,700)                  6      0.24
 [700,900)                  8      0.32
 [900,1.2e+03)              2      0.08
 [1.2e+03,1.45e+03]         3      0.12
> sum<-c(0,cumsum(tab$Rel.Freq))
> sum
[1] 0.00 0.24 0.48 0.80 0.88 1.00
> par(mfrow=c(1,2))
> hist(Income,breaks=class,freq=T,include.lowest=T,right=F)
> plot(class,sum,type="b",xlab="Income",main="Empirical cdf")
```



8. By using R:

```
> ecdf(Income)(1300)-ecdf(Income)(750)
[1] 0.36
```

36% of students has income from 750 to 1300€.

```
> 1-ecdf(Income)(800)
[1] 0.32
```

32% of students has income more than 800€.

```
> quantile(Income,0.5)
50%
700
```

The highest income of the 50% of the students with the lowest income is equal to 700€.

```
> quantile(Income,0.8)
80%
864.8
```

The smallest income of the 20% of the students with the highest income is equal to 864.80€.