M1 MPM - Dunkerque, Pôle Lamartine - 2013/2014

CORRECTION Exercices Chapitre 1 - Statistiques à une variable.

Exercice 1

1. La population étudiée est constituée de l'ensemble des patients du fichier du service ORL du centre hospitalier de Dunkerque.
   Les individus de la population sont les patients.
2. (a) Sexe : variable qualitative. Modalités : F,M.
   (b) Âge : variable quantitative discrète ou continue. Modalités (en années) : $20, 21, 22, \ldots, [20; 25[, [25; 30[, \ldots$
   (c) Profession : variable qualitative. Modalités : ouvrier, artisan,...
   (d) Poids : variable quantitative discrète ou continue. Modalités (en kg) : $65, 70, 75, \ldots, [60; 65[, [65; 70[, \ldots$
   (e) Taille : variable quantitative discrète ou continue. Modalités (en cm) : $160, 165, 170, \ldots, [160; 165[, [165; 170[, \ldots$
   (f) Groupe sanguin : variable qualitative. Modalits : 0^+,A^-,...

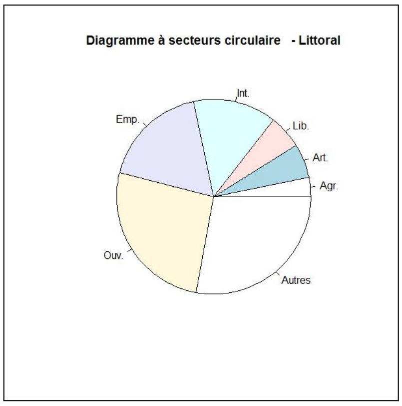
Exercice 2

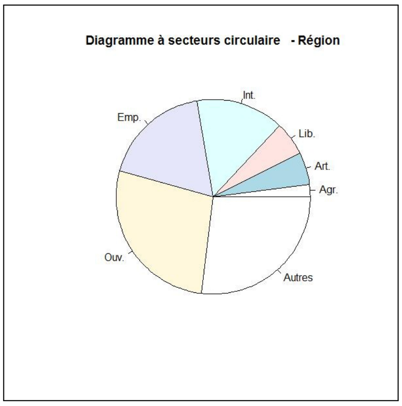
1. La population étudiée est constituée de l'ensemble des communes françaises de plus de 20000 habitants.
   Les individus de la population sont les communes.
2. (a) Département : variable qualitative (même si le département est décrit par un nombre, celui n'est pas quantitativement utilisable). Modalités : 59,62,...
   (b) Nombre d'habitants : variable quantitative continue. Modalités : $[20000; 25000[, [25000; 30000[, \ldots$
   (c) Nombre d'établissements d'enseignement secondaire : variable quantitative continue. Modalités : $[5; 10[, [10; 15[, \ldots$

Exercice 3

1. La variable statistique étudiée est la CSP (Catégorie Socio-Professionnelle) des actifs du Littoral et de la région Nord-Pas de Calais. Cette variable est une variable qualitative.
2. Les représentations graphiques envisageables sont le diagramme à secteurs circulaire, le diagramme en tuyaux d'orgue et le diagramme en bandes.
3. On donne le diagramme à secteurs circulaire, généré sous `R` par les commandes suivantes :

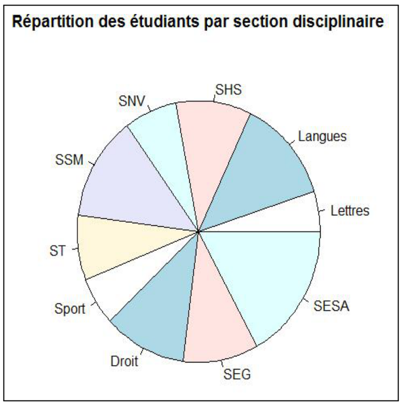
```
> Littoral<-c(3.2,5.7,5.6,13.8,17.7,26.1,27.9)
> Region<-c(2,5.3,5.6,14.4,17.8,27,26.6)
> pie(Littoral,main='Diagramme à secteurs circulaire - Littoral',
+ labels=c('Agr.','Art.','Lib.','Int.','Emp.','Ouv.','Autres'))
> pie(Region,main='Diagramme à secteurs circulaire - Région'
+ labels=c('Agr.','Art.','Lib.','Int.','Emp.','Ouv.','Autres'))
```





Exercice 4 On donne le diagramme à sections circulaire, généré sous `R` par les commandes suivantes :

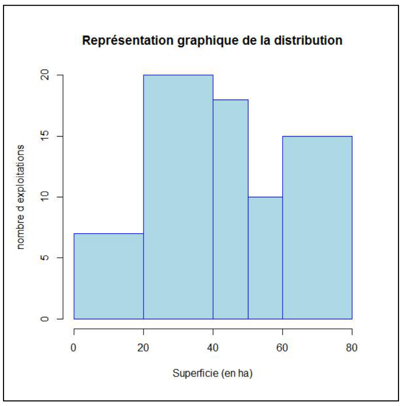
```
> rep.etu<-c(5,13,10,7,13,8,6,11,10,17)/100
> pie(rep.etu,main='Répartition des étudiants par section disciplinaire',
+ labels=c('Lettres','Langues','SHS','SNV','SSM','ST','Sport','Droit','SEG','SESA'))
```



Exercice 5

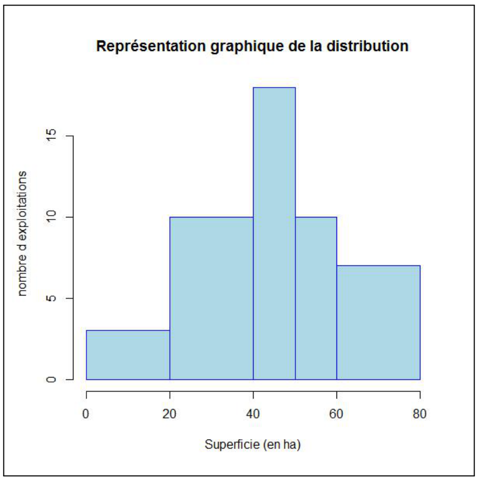
1. La variable étudiée est la superficie en hectares des exploitations agricoles de la commune. Cette variable est quantitative continue.
2. On utilise pour représenter graphiquement cette distribution l'histogramme obtenu sous `R` à l'aide des commandes suivantes :

```
> classes<-c(0,20,40,50,60,80)
> centres.classes<-c(10,30,45,55,70)
> exploitations<-c(7,20,18,10,15)
> serie<-c(rep(centres.classes,exploitations))
> # Il est nécessaire de créer une suite de 70 valeurs appelée ici "serie" qui respecte la
> # distribution proposée (et les amplitudes de classes) afin de pouvoir utiliser la fonction
hist.
> hist(serie,classes,xlab='Superficie (en ha)',ylab='nombre d exploitations',
+ main='Représentation graphique de la distribution',col='lightblue',border='blue',freq=T)
```



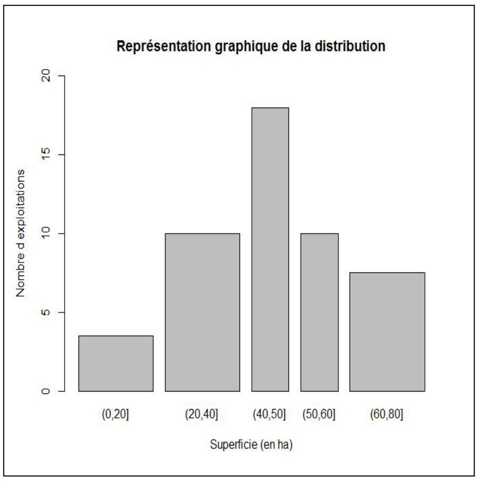
On remarque cependant que `R` ne prend pas en compte les différentes amplitudes de classes. Il faut donc le faire à sa place.

```
> ampl<-classes[2 :length(classes)]-classes[1 :length(classes)-1]
> ampl.ref<-min(ampl)
> coeff<-ampl/ampl.ref
> exploitationsOK<-exploitations/coeff
> serieOK<-c(rep(centres.classes,exploitationsOK))
> hist(serieOK,classes,xlab='Superficie (en ha)',ylab='nombre d exploitations',
+ main='Représentation graphique de la distribution',col='lightblue',border='blue',freq=T)
```



On peut également utiliser la fameuse fonction `barplot` :

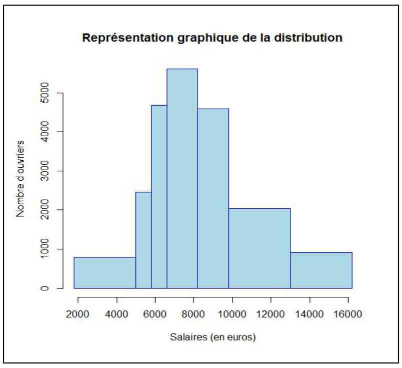
```
> tab<-table(cut(serie,c(0,20,40,50,60,80)))
> tabOK<-tab/coeff
> barplot(tabOK,ylim=c(0,20),width=ampl)
```



Exercice 6

1. La variable étudiée est le salaire en euros des personnels ouvriers du groupe industriel considéré. Cette variable est quantitative continue.
   L'information relative à la masse salariale des classes extrêmes permet de préciser ces classes ainsi que leur centre :
   - La masse des salaires correspondant à la première classe (moins de 5000 euros) s'élève à 10,693 millions d'euros. Comme cette classe compte 3145 ouvriers, on en déduit que le centre de classe vaut : $\frac{10693000}{3145} = 3400$ et la classe associée est donc $[1800; 5000[$.
   - La masse des salaires correspondant à la dernière classe s'élève à 53,363 millions d'euros. Comme cette classe compte 3655 ouvriers, on en déduit que le centre de classe vaut : $\frac{53363000}{3655} = 14600$ et la classe associée est $[13000; 16200[$.
2. On utilise pour représenter graphiquement cette distribution l'histogramme obtenu sous `R` à l'aide des commandes suivantes :

```
> classes<-c(1.8,5,5.8,6.6,8.2,9.8,13,16.2)*1000
> centres.classes<-c(3.4,5.4,6.2,7.4,9,11.4,14.6)*1000
> ouvriers<-c(3145,2465,4675,11220,9180,8160,3655)
> ampl<-classes[2 :length(classes)]-classes[1 :length(classes)-1]
> ampl.ref<-min(ampl)
> coeff<-ampl/ampl.ref
> ouvriersOK<-ouvriers/coeff
> serieOK<-c(rep(centres.classes,ouvriersOK))
> hist(serieOK,breaks=classes,main='Représentation graphique de la distribution',
+ xlab='Salaires (en euros)',ylab='Nombre d ouvriers',col='lightblue',border='blue',freq=T)
```



Exercice 7

1. Les modalités sont peu nombreuses et toutes distinctes, il est donc inutile de dresser un tableau synthétique. On a $\overline{x} = \frac{1+5+4+3+7+6+10}{7} = 5,14$ à 10^{-2} près par défaut. Le nombre d'individus est égal à $n = 7 = 2 \times 3 + 1$. Une fois la série ordonnée par ordre de grandeur croissant ou décroissant, on trouve facilement Mé=5. Sous R, on tape :

```
> erreurs<-c(1,5,4,3,7,6,10)
> mean(erreurs)
[1] 5.142857
> median(erreurs)
[1] 5
```

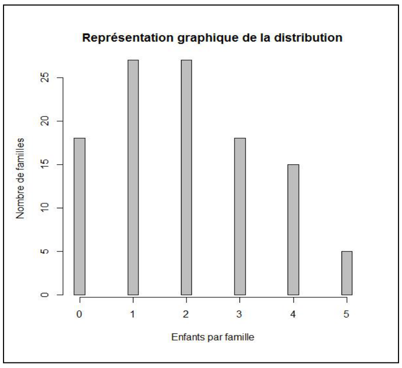
2. Les modalités sont toujours peu nombreuses et distinctes. La moyenne devient $\overline{x} = \frac{1+5+4+3+7+6+10+9}{8} = 5,62$ à 10^{-2} près par défaut. Le nombre d'individus de la population passe à $n = 8 = 2 \times 4$. La série étant ordonnée, on obtient Mé= $\frac{5+6}{2} = 5,5$. Sous R, on tape :

```
> nouv.erreurs<-c(erreurs,9)
> mean(nouv.erreurs)
[1] 5.625
> median(nouv.erreurs)
[1] 5.5
```

Exercice 8

1. La variable étudiée (nombre d'enfants par famille) est décrite dans le tableau comme une variable (quantitative) discrète. La distribution sera donc représentée graphiquement à l'aide d'un diagramme en bâtons. On utilise le code suivant sous R :

```
> enfants<-c(0,1,2,3,4,5)
> familles<-c(18,27,27,18,15,5)
> diag<-tapply(familles,enfants,sum)
> barplot(diag,space=4,main='Représentation graphique de la distribution',
+ xlab='Enfants par famille',ylab='Nombre de familles',axis.lty=1)
```



2. On complète le tableau de l'énoncé afin de répondre à cette question et à la suivante :

Nombre d'enfants x_i	Nombre de familles n_i	$n_i \nearrow$	$n_i x_i$	$n_i x_i^2$
0	18	18	0	0
1	27	45	27	27
2	27	72	54	108
3	18	90	54	162
4	15	105	60	240
5	5	110	25	125
TOTAL	110	–	220	662

Le nombre moyen d'enfants de ces familles vaut $\overline{x} = \frac{220}{110} = 2$.

3. • On a $n = 110 = 2 \times 55$. Par conséquent, Mé$= \frac{2+2}{2} = 2$.
   • La série est bimodale (deux modalités sont associées à l'effectif le plus élevé). Les modes sont 1 et 2.
   • L'écart-type vaut $\sigma(X) = \left(\frac{662}{110} - 2^2\right)^{\frac{1}{2}} = 1,42$ à 10^{-2} près par défaut.

   Retrouvons ces derniers résultats à l'aide de `R` :

```
> serie<-c(rep(enfants,familles))
> mean(serie)
[1] 2
> median(serie)
[1] 2
> dat<-data.frame(enfants,familles)
> dat$enfants[which(dat$familles==max(dat$familles))]    # Donne le mode
[1] 1 2
> sd(serie)
[1] 1.427129
> # On remarque que les résultats ne coïncident pas. C'est normal car
> # la fonction sd calcule l'écart-type corrigé de la série. Il faut donc
> # modifier le calcul de la manière suivante :
> n<-110
> sd(serie)*sqrt((n-1)/n)
[1] 1.420627
```

4. La proportion de familles comptant au plus 3 enfants est égale à $p = \frac{18+27+27+18}{110} \times 100 = 81,82\%$ à 10^{-2} près par excès.
5. La proportion de familles comptant au moins 3 enfants est égale à $p = \frac{18+15+5}{110} \times 100 = 34,55\%$ à 10^{-2} près par excès.

Exercice 9 On se donne le tableau suivant :

Superficie (en ha)	Centre de classe x_i	Nombre d'exploitations n_i	$n_i \nearrow$	$n_i x_i$	$n_i x_i^2$
0 à moins de 20	10	7	7	70	700
20 à moins de 40	30	20	27	600	18000
40 à moins de 50	45	18	45	810	36450
50 à moins de 60	55	10	55	550	30250
60 à moins de 80	70	15	70	1050	73500
TOTAL	–	70	–	3080	158900

• La moyenne vaut $\overline{x} = \frac{3080}{70} = 44$.
• La médiane appartient à la classe $[40; 50[$ car $\frac{n}{2} = \frac{70}{2} = 35$. On utilise l'interpolation linéaire afin d'être plus précis. On a les valeurs suivantes :

40	Mé	50
27	$\frac{n}{2} = 35$	45

ce qui permet d'affirmer que Mé$= 40 + \frac{(35-27)}{(45-27)} \times (50-40) = 44,44$ à 10^{-2} près par défaut.
• On procède de même pour les quartiles.

20	Q_1	40
7	$\frac{n}{4}=17,5$	27

$\Rightarrow \quad Q_1 = 20 + \frac{(17,5-7)}{(27-7)} \times (40-20) = 30,5$ à 10^{-2} près par défaut.

50	Q_3	60
45	$\frac{3n}{4}=52,5$	55

$\Rightarrow \quad Q_3 = 50 + \frac{(52,5-45)}{(55-45)} \times (60-50) = 57,5.$

- L'écart-type vaut $\sigma(X) = \left(\frac{73500}{70} - 44^2\right)^{\frac{1}{2}} = 18,28$ à 10^{-2} près par excès.
- La classe modale est la classe $[20;40[$ car c'est elle qui implique l'effectif le plus important.
- La proportion d'exploitations agricoles dont la superficie est inférieure à 45 hectares ne peut pas être calculée immédiatement. Déterminons l'effectif théorique des exploitations ayant une superficie en hectares comprise entre 40 et 45 hectares (et non entre 40 et 50 hectares ce qui nous aurait facilité la tâche...), toujours grâce à l'interpolation linéaire :

40	45	50
0	p_0	18

$\Rightarrow \quad p_0 = 0 + \frac{(45-40)}{(50-40)}(18-0) = 9$ (ce qui somme toute est très logique). Par conséquent la proportion recherchée est égale à $\frac{7+20+9}{70} \times 100 = 51,43$ à 10^{-2} près par excès.

Retrouvons quelques-unes de ces valeurs grâce à `R` :

```
> centres.classes<-c(10,30,45,55,70)
> exploitations<-c(7,20,18,10,15)
> serie<-c(rep(centres.classes,exploitations))
> mean(serie)
[1] 44
> median(serie)
[1] 45
> # R traite la distribution de valeurs comme une série discrète d'où les différences
> quantiles(serie)
 0%  25%  50%  75%  100%
 10   30   45   55    70
> sd(serie)*sqrt((70-1)/70)
[1] 18.27567
> dat<-data.frame(centres.classes,exploitations)
> dat$centres.classes[which(dat$exploitations==max(dat$exploitations))]
[1] 30
> # On en déduit que la classe modale est la classe [20;40[
```

Exercice 10 On se donne le tableau suivant :

Salaire annuel	Centre de classe x_i	Nombre d'ouvriers n_i	$n_i \nearrow$	$n_i x_i$	$n_i x_i^2$
moins de 5000 euros	3400	3145	3145	10693000	3,635620e+10
5000 à moins de 5800 euros	5400	2465	5610	13311000	7,187940e+10
5800 à moins de 6600 euros	6200	4675	10285	28985000	1,797070e+11
6600 à moins de 8200 euros	7400	11220	21505	83028000	6,144072e+11
8200 à moins de 9800 euros	9000	9180	30685	82620000	7,435800e+11
9800 à moins de 13000 euros	11400	8160	38845	93024000	1,060474e+12
13000 euros et plus	14600	3655	42500	53363000	7,790998e+11
TOTAL	–	42500	–	365024000	3,485503e+12

- La moyenne vaut $\overline{x} = \frac{365024000}{42500} = 8588,8$.
- La médiane appartient à la classe $[6600;8200[$ car $\frac{n}{2} = \frac{42500}{2} = 21250$. On utilise l'interpolation linéaire afin d'être plus précis. On a les valeurs suivantes :

6600	Mé	8200
10285	$\frac{n}{2}=21250$	21505

ce qui permet d'affirmer que Mé$= 6600 + \frac{(21250-10285)}{(21505-10285)} \times (8200-6600) = 8163,64$ à 10^{-2} près par excès.

- L'écart-type vaut $\sigma(X) = \left(\frac{3,485503e+12}{42500} - 8588,8^2\right)^{\frac{1}{2}} = 2871,30$ à 10^{-2} près par excès.
- La classe modale est la classe $[6600;8200[$ car c'est elle qui implique l'effectif le plus important.

Retrouvons quelques-unes de ces valeurs grâce à `R` :

```
> centres.classes<-c(3400,5400,6200,7400,9000,11400,14600)
> ouvriers<-c(3145,2465,4675,11220,9180,8160,3655)
> serie<-c(rep(centres.classes,ouvriers))
> median(serie)
[1] 7400
> # Comme précédemment, on n'obtient pas la valeur correcte de la médiane.
> mean(serie)
[1] 8588.8
> sd(serie)*sqrt((length(serie)-1)/length(serie))
[1] 2871.298 > dat<-data.frame(centres.classes,ouvriers)
> dat$centres.classes[which(dat$ouvriers==max(dat$ouvriers))]
[1] 7400
> # On en déduit que la classe modale est la classe [6600;8200[
```

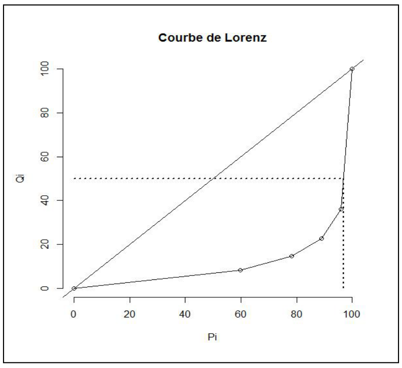
Exercice 11

1. On complète le tableau de l'énoncé :

Nombre de salariés	Nombre entreprises $n_i(\%)$	$n_i \nearrow = P_i(\%)$	Chiffre d'affaires	Chiffre d'affaire $\nearrow = Q_i(\%)$
$[20;50[$	59,7	59,7	8,3	8,3
$[50;100[$	18,5	78,2	6,2	14,5
$[100;200[$	10,9	89,1	8,1	22,6
$[200;500[$	7	96,1	13,4	36
500 et plus	3,9	100	64	100

On utilise le code suivant sous R :

```
> pi<-c(59.7,18.5,10.9,7,3.9)
> qi<-c(8.3,6.2,8.1,13.4,64)
> Pi<-c(0,cumsum(pi))
> Qi<-c(0,cumsum(qi))
> plot(Pi,Qi,xlim=c(0,100),ylim=c(0,100),main='Courbe de Lorenz',xlab='Pi',ylab='Qi',
+ bty='n',xaxt='n',yaxt='n')
> lines(Pi,Qi)
> abline(h=0)
> abline(v=0)
> abline(0,1)
> segments(0,50,50,50,lty='dotted')
> segments(50,50,50,0,lty='dotted')
```



2. On trouve à l'aide du graphique précédent que la part des entreprises (ayant le moins de salariés) qui réalisent 50% du chiffre d'affaires total de ce secteur est approximativement 97%, ce qui indique une distribution très inégalitaire du chiffre d'affaires.
3. Par définition, $i = \frac{S}{S_0}$ où
   - S_0 désigne l'aire située entre la première bissectrice, l'axe des abscisses et la droite d'équation $Pi = 100$ donc $S_0 = 5000$,
   - S désigne l'aire située entre la première bissectrice, la courbe de Lorenz et la droite d'équation $Pi = 100$. On a $S =$

$S_0 - \left(\frac{(59,7-0)\times(8,3+0)}{2} + \frac{(78,2-59,7)(14,5+8,3)}{2} + \frac{(89,1-78,2)(22,6+14,5)}{2} + \frac{(96,1-89,1)(36+22,6)}{2} + \frac{(100-96,1)(100+36)}{2}\right)$
$= 5000 - (247,76 + 210,9 + 202,20 + 205,1 + 265,2) = 5000 - 1131,16 = 3868,84.$
On en déduit que $i = \frac{3868,84}{5000} = 0,77$ (à 10^{-2} près par défaut), ce qui indique une concentration forte du chiffre d'affaires, ce qui corrobore la remarque de la question précédente.

4. D'après le tableau, la classe médiane est $[20;50[$ car elle implique 50% de l'effectif total. Déterminons la médiane à l'aide de l'interpolation linéaire. On a les valeurs suivantes :

20	Mé	50
0	$\frac{n}{2} = 50$	59,7

ce qui permet d'affirmer que Mé$= 20 + \frac{(50-0)}{(59,7-0)} \times (50-20) = 45,12$ à 10^{-2} près par défaut. Par conséquent, 50% des entreprises de l'étude ont un nombre de salariés inférieur à $45,12\%$.

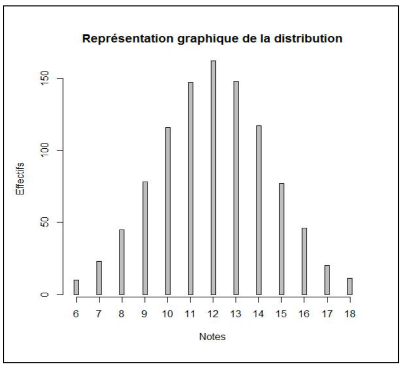
Exercice 12

1. On a le tableau :

Notes x_i	Effectifs n_i	Fréquences f_i	$n_i \nearrow$	Fcc $f_i \nearrow$
6	10	1	10	60
7	23	2,3	33	221
8	45	4,5	78	581
9	78	7,8	156	15,6
10	116	11,6	272	27,2
11	147	14,7	419	41,9
12	162	16,2	581	58,1
13	148	14,8	729	72,9
14	117	11,7	846	84,6
15	77	7,7	923	92,3
16	46	4,6	969	96,9
17	20	2	989	98,9
18	11	1,1	1000	100
TOTAL	1000	100	–	–

2. On utilise le code suivant sous R :

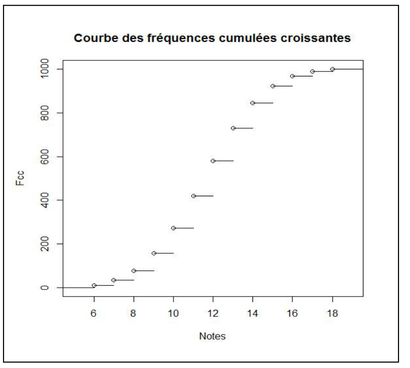
```
> Notes<-seq(6,18,1)
> Eff<-c(10,23,45,78,116,147,162,148,117,77,46,20,11)
> diag<-tapply(Eff,Notes,sum)
> barplot(diag,space=4,main='Représentation graphique de la distribution',
+ xlab='Notes',ylab='Effectifs',axis.lty=1)
```



3. On utilise le code suivant sous R :

```
> Effcum<-cumsum(Eff)
```

```
> diagesc<-stepfun(Notes,c(0,Effcum))
> plot(diagesc,vertical=FALSE,main='Courbe des fréquences cumulées croissantes',
> xlab='Notes',ylab='Fcc')
```



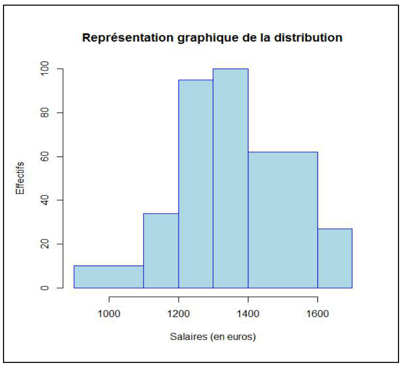
Exercice 13

1. Le tableau une fois rempli est donné par :

Salaire en euros	Centres	Effectifs	Fréquences (%)	Effectifs cumulés croissants
$[900; 1100[$	1000	20	5	20
$[1100; 1200[$	1150	34	8,5	54
$[1200; 1300[$	1250	95	23,75	149
$[1300; 1400[$	1350	100	25	249
$[1400; 1600[$	1500	124	31	373
$[1600; 1700[$	1650	27	6,75	400
TOTAL	–	400	100	–

2. On dénombre $20 + 34 + 95 = 149$ employés ayant un salaire inférieur à 1300 euros.
3. On utilise le code :

```
> classes<-c(900,1100,1200,1300,1400,1600,1700)
> centres.classes<-c(1000,1150,1250,1350,1500,1650)
> effectifs<-c(20,34,95,100,124,27)
> ampl<-classes[2 :length(classes)]-classes[1-length(classes)-1]
> ampl.ref<-min(ampl)
> coeff<-ampl/ampl.ref
> effectifsOK<-effectifs/coeff
> serieOK<-c(rep(centres.classes,effectifsOK))
> hist(serieOK,classes,xlab='Salaires (en euros)',ylab='Effectifs',
+ main='Représentation graphique de la distribution',col='lightblue',border='blue',freq=T)
```

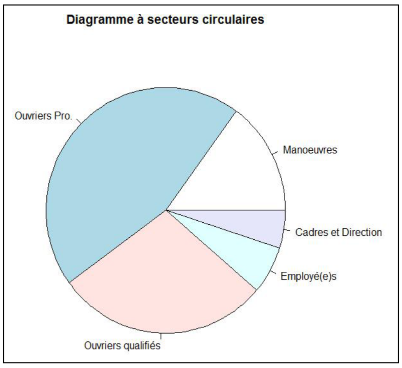


Exercice 14

1. On complète le tableau de l'énoncé :

Fonction	Nombre	Pourcentage	Part (en degrés)
Manœuvres	96	15	54
Ouvriers professionnels	288	45	162
Ouvriers qualifiés	184	28,75	103,5
Employé(s)s	40	6,25	22,5
Cadres et Direction	32	5	18
TOTAL	640	100	360

2. On a le diagramme à secteurs circulaire suivant :



obtenu grâce au code `R` :

```
> Personnel<-c(96,288,184,40,32)
> lbls<-c('Manoeuvres','Ouvriers Pro.','Ouvriers qualifiés','Employé(e)s','Cadres et Direction')
> pie(Personnel,labels,main='Diagramme à secteurs circulaires')
```

Exercice 15

1. Ordonnons les valeurs de série par ordre de grandeur croissant :

$$3, 3, 4, 4, 4, 5, 6, 7, 7, 8, 9, 9, 10$$

Comme $n = 13 = 2 \times 6 + 1$, la médiane est la valeur située à la 7e position. On a donc Mé= 6.

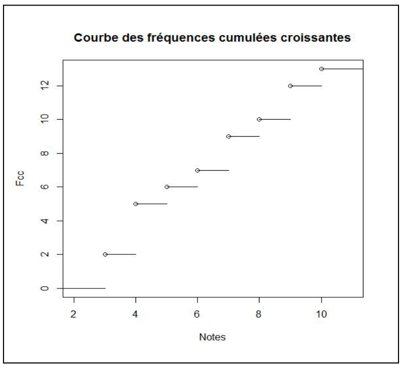
2. On a le tableau de distribution suivant :

x_i	n_i	$n_i \nearrow$
3	2	2
4	3	5
5	1	6
6	1	7
7	2	9
8	1	10
9	2	12
10	1	13

3. On cherche $k + 1 = 7$ dans la colonne des effectifs cumulés croissants, qui apparaît explicitement à la 4e ligne du tableau, ce qui nous permet de retrouver Mé= 6.

4. On utilise le code suivant sous `R` :

```
> valeurs<-c(3,4,5,6,7,8,9,10)
> Eff<-c(2,3,1,1,2,1,2,1)
> Effcum<-cumsum(Eff)
> diagesc<-stepfun(valeurs,c(0,Effcum))
> plot(diagesc,vertical=FALSE,main='Courbe des effectifs cumulés croissants',
> xlab='Notes',ylab='Fcc')
```



Exercice 16

1. On a le tableau :

Nombre x_i de pièces A	Effectif n_i	$n_i \nearrow$	$n_i x_i$	$n_i x_i^2$
6	1	1	6	36
7	3	4	21	147
8	3	7	24	192
9	5	12	45	405
10	2	14	20	200
11	2	16	22	242
12	4	20	48	576
13	2	22	26	338
14	5	27	70	980
15	5	32	75	1125
16	3	35	48	768
17	1	36	17	289
18	3	39	54	972
20	1	40	20	400
TOTAL	40	–	496	6670

2. Comme $n = 40 = 2 \times 20$, la médiane est la 1/2-somme des $k = 20e$ et $k+1 = 21e$ valeurs de la série ordonnée par ordre de grandeur croissant. À l'aide du tableau, on obtient Mé$= \frac{12+13}{2} = 12,5$.
3. $\overline{x} = \frac{496}{40} = 12,4$.
4. $V(x) = \frac{6670}{40} - (12,4)^2 = 12,99$. On en déduit $\sigma(X) = \sqrt{12,99} = 3,60$ à 10^{-2} près. On en conclut que l'intervalle de dispersion est donné par
$$[\overline{x} - \sigma(X); \overline{x} + \sigma(X)] = [12,99 - 3,60; 12,99 + 3,60] = [9,39; 16,59].$$
On peut alors supposer que 68% des valeurs de la série sont théoriquement dans cet intervalle.

Retrouvons ces valeurs à l'aide de `R` :

```
> n<-40
> Pieces<-c(seq(1,18,1),20)
> Eff<-c(1,3,3,5,2,2,4,2,5,5,3,1,3,1)
> serie<-c(rep(Pieces,Eff))
> median(serie)
[1] 12.5
> mean(serie)
[1] 12.4
> var(serie)*(n-1)/n
[1] 12.99
> sd(serie)*sqrt((n-1)/n)
[1] 3.604164
```

Exercice 17 Complétons le tableau :

Production journalière de véhicules équipés x_i	Nombre de jours n_i	$n_i x_i$	$n_i x_i^2$
95	1	95	9025
96	3	288	27648
97	6	582	56454
98	8	784	76832
99	10	990	98010
100	13	1300	130000
101	18	1818	183618
102	14	1428	145656
103	9	927	95481
104	8	832	86528
105	6	630	66150
106	2	212	22472
107	2	214	22898
TOTAL	100	10100	1020772

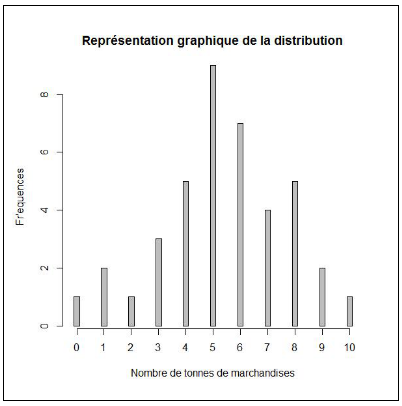
On a donc $\overline{x} = \frac{10100}{100} = 101$ et $\sigma(X) = \left(\frac{1020772}{100} - 101^2\right)^{\frac{1}{2}} = \sqrt{6,72} = 2,59$ à 10^{-2} près.
On obtient les mêmes résultats avec `R` :

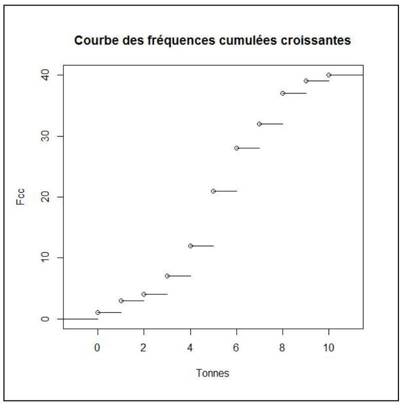
```
> n<-100
> prod<-seq(95,107,1)
> eff<-c(1,3,6,8,10,13,18,14,9,8,6,2,2)
> serie<-rep(prod,eff)
> mean(serie)
[1] 101
> sd(serie)*sqrt((n-1)/n)
[1] 2.592296
```

Exercice 18 On dresse le tableau suivant :

Tonnes x_i	Effectifs n_i	$n_i \nearrow$
0	1	1
1	2	3
2	1	4
3	3	7
4	5	12
5	9	21
6	7	28
7	4	32
8	5	37
9	2	39
10	1	40
TOTAL	40	–

On obtient les deux graphiques ci-dessous :





grâce aux codes R ci-dessous :

```
> Tonnes<-0 :10
> Eff<-c(1,2,1,3,5,9,7,4,5,2,1)
> diag<-tapply(Eff,Tonnes,sum)
> barplot(diag,space=4,main='Représentation graphique de la distribution',
+ xlab='Nombre de tonnes de marchandises',ylab='Fréquences',axis.lty=1)

> Effcum<-cumsum(Eff)
> diagesc<-stepfun(Tonnes,c(0,Effcum))
> plot(diagesc,vertical=FALSE,main='Courbe des effectifs cumulés croissants',
> xlab='Tonnes',ylab='Fcc')
```

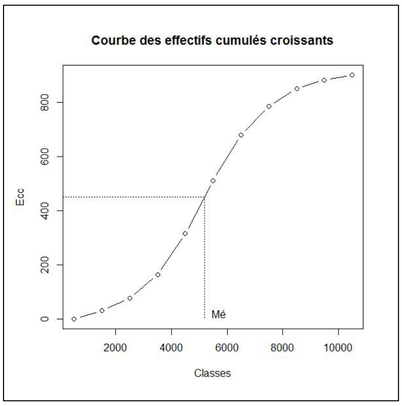
Exercice 19

1.

i	Classes	Centres x_i	Effectifs n_i	$n_i \nearrow$
1	$500 \leq X < 1500$	1000	31	31
2	$1500 \leq X < 2500$	2000	46	77
3	$2500 \leq X < 3500$	3000	86	163
4	$3500 \leq X < 4500$	4000	151	314
5	$4500 \leq X < 5500$	5000	197	511
6	$5500 \leq X < 6500$	6000	167	678
7	$6500 \leq X < 7500$	7000	107	785
8	$7500 \leq X < 8500$	8000	65	850
9	$8500 \leq X < 9500$	9000	32	882
10	$9500 \leq X < 10500$	10000	18	900
TOTAL	–	–	900	–

2. On utilise le code suivant afin de générer la courbe polygonale attendue :

```
> Classes<-c(500,1500,2500,3500,4500,5500,6500,7500,8500,9500,10500)
> Eff<-c(0,31,46,86,151,197,167,107,65,32,18)
> Effcum<-cumsum(Eff)
> plot(Classes,Effcum,type='b',main='Courbe des effectifs cumulés croissants',
+ xlab='Classes',ylab='Ecc')
> segments(0,450,5190,450,lty='dotted')
> segments(5190,450,5190,0,lty='dotted')
> text(5700,20,"Mé")
```



On trouve approximativement Mé= 5200.

3. On utilise l'interpolation linéaire et donc les valeurs suivantes :

4500	Mé	5500
314	$\frac{n}{2} = 450$	511

$\Rightarrow$ Mé$= 4500 + \frac{(450-314)}{(511-314)} \times (5500 - 4500) = 5190,355$ à 10^{-3} près par défaut.

Exercice 20

1. (a) On se donne le tableau :

Distance (en km)	Centres de classe x_i	Effectifs n_i	$n_i \nearrow$	$n_i x_i$
$[0;5[$	2,5	50	50	125
$[5;10[$	7,5	250	300	1875
$[10;15[$	12,5	500	800	6250
$[15;20[$	17,5	800	1600	14000
$[20;25[$	22,5	700	2300	15750
$[25;30[$	27,5	650	2950	17875
$[30;35[$	32,5	320	3270	10400
$[35;40[$	37,5	230	3500	8625
TOTAL	–	3500	–	74900

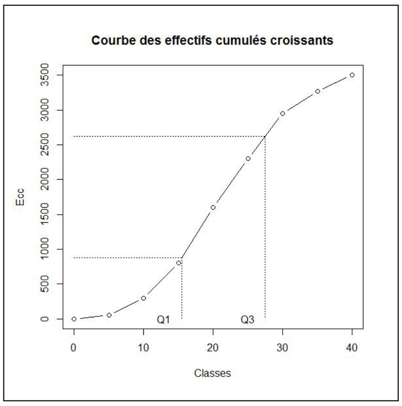
Le nombre de kilomètres parcourus pour chaque classe est donné dans la dernière colonne. Le nombre total de kilomètres parcourus est égal à $\sum_{i=1}^{p} n_i x_i = 74900$.

(b) Le prix de revient étant de 5 par kilomètres, le coût global est égal à $74900 \times 5 = 374500$. Comme chaque livraison est facturée 100 et qu'on dénombre 3500 livraisons, la recette est égale à $3500 \times 100 = 350000$. Le service est bien déficitaire et la perte est estimée à $374500 - 350000 = 24500$.

2. (a) Voir le tableau précédent.

(b) On utilise le code :

```
> Classes<-c(0,5,10,15,20,25,30,35,40)
> Effectifs<-c(0,50,250,500,800,700,650,320,230)
> Effcum<-cumsum(Effectifs)
> plot(Classes,Effcum,type='b',main='Courbe des effectifs cumulés croissants',
+ xlab='Classes',ylab='Ecc')
> segments(0,875,15.5,875,lty='dotted')
> segments(15.5,875,15.5,0,lty='dotted')
> text(13,0,'Q1')
> segments(0,2625,27.5,2625,lty='dotted')
> segments(27.5,2625,27.5,0,lty='dotted')
> text(25,0,'Q3')
```



$$\begin{array}{c|c|c} 800 & \frac{n}{4}=875 & 1600 \\ \hline 15 & Q_1 & 20 \end{array} \Rightarrow Q_1 = \frac{(875-800)}{(1600-800)} \times (20-15)+15 = 15,47 \text{ à } 10^{-2} \text{ près.}$$

$$\begin{array}{c|c|c} 2300 & \frac{3n}{4}=2625 & 2950 \\ \hline 25 & Q_3 & 30 \end{array} \Rightarrow Q_3 = \frac{(2625-2300)}{(2950-2300)} \times (30-25)+25 = 27,5.$$

(c) L'intervalle interquartile est $[15,47;27,5]$ et l'interquartile est égal à $Q_3 - Q_1 = 27,5 - 15,47 = 12,03$.

3. (a)

Classes	Effectifs	kilomètres parcourus	Coût des livraisons	Coût moyen d'une livraison
$[0;15[$	800	6000	30000	37,5
$[15;30[$	2150	48375	241875	112,5
$[30;40[$	550	19250	96250	175

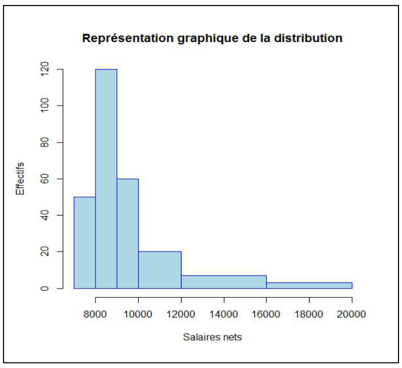
(b) On peut conseiller au gérant de se concentrer sur les livraisons situées à moins de 15km du magasin.

Exercice 21

1. Le tableau est le suivant :

Classes	x_i	Effectifs n_i	$n_i \nearrow$	$n_i \searrow$	$n_i x_i (\times 10^3)$	$n_i x_i^2 (\times 10^6)$	u_i	$n_i u_i$	$n_i u_i^2$
$[7000;8000[$	7500	50	50	315	375	2812,5	-4	-200	800
$[8000;9000[$	8500	120	170	265	1020	8670	-2	-240	480
$[9000;10000[$	9500	60	230	145	570	5415	0	0	0
$[10000;12000[$	11000	40	270	85	440	4840	3	120	360
$[12000;16000[$	14000	30	300	45	420	5880	9	270	2430
$[16000;20000[$	18000	15	315	15	270	4860	17	255	4355
TOTAL	–	315	–	–	3095	32477,5	–	205	8405

2. On a l'histogramme



grâce au code R suivant :

```
> n<-315
```

```
> classes<-c(7000,8000,9000,10000,12000,16000,20000)
> centres.classes<-c(7500,8500,9500,11000,14000,18000)
> effectifs<-c(50,120,60,40,30,15)
> serie<-rep(centres.classes,effectifs)
> ampl<-classes[2 :length(classes)]-classes[1 :length(classes)-1]
> ampl.ref<-min(ampl)
> coeff<-ampl/ampl.ref
> effectifsOK<-effectifs/coeff
> serieOK<-c(rep(centres.classes,effectifsOK))
> hist(serieOK,classes,xlab='Salaires nets',ylab='Effectifs',
+ main='Représentation graphique de la distribution',col='lightblue',border='blue',freq=T)
```

3. L'étendue est égale à $e = 20000 - 7000 = 13000$.
4. Comme $\frac{n}{2} = \frac{315}{2} = 157,5$, la 3e colonne du tableau nous permet d'affirmer que la classe modale est $[8000; 9000[$.
5. $\begin{array}{c|c|c} 50 & \frac{n}{4} = 78,75 & 170 \\ \hline 8000 & Q_1 & 9000 \end{array} \Rightarrow Q_1 = 8000 + \frac{(78,75-50)}{(170-50)}(9000-8000) = 8239,58$ à 10^{-2} près,
   $\begin{array}{c|c|c} 50 & \frac{n}{2} = 157,5 & 170 \\ \hline 8000 & Q_2 & 9000 \end{array} \Rightarrow Q_2 = 8000 + \frac{(157,5-50)}{(170-50)}(9000-8000) = 8895,83$ à 10^{-2} près,
   $\begin{array}{c|c|c} 230 & \frac{3n}{4} = 236,25 & 270 \\ \hline 10000 & Q_3 & 12000 \end{array} \Rightarrow Q_3 = 10000 + \frac{(236,25-230)}{(270-230)}(12000-10000) = 10312,5.$
6. • $\overline{x} = \frac{3095000}{315} = 9825,40$ à 10^{-2} près par excès,
   $\sigma(X) = \left(\frac{32477,5\times 10^6}{315} - (9825,40)^2\right)^{\frac{1}{2}} = 2562,18$ à 10^{-2} près par excès.
   • $\overline{u} = \frac{205}{315} = 0,6507937$ et $\overline{x} = \overline{u} \times 500 + 9500 = 9825,40$ à 10^{-2} près par excès,
   $\sigma(U) = \left(\frac{8405}{315} - (0,6507937)^2\right)^{\frac{1}{2}} = 5,124354$ et $\sigma(X) = 500 \times \sigma(U) = 2562,18$ à 10^{-2} près par excès.
   Avec R :

```
> mean(serie)
[1] 9825.397
> sd(serie)*sqrt((n-1)/n)
[1] 2562.177
```

Exercice 22

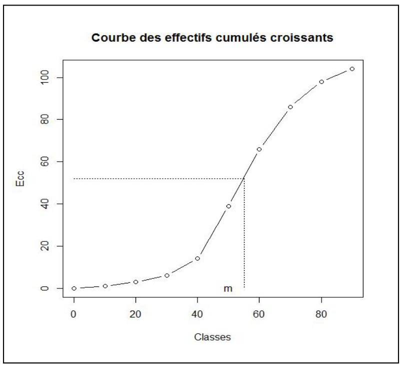
1.

Unités (×100)	Nombre de semaines n_i	$n_i \nearrow$
0 à moins de 10	1	1
10 à moins de 20	2	3
20 à moins de 30	3	6
30 à moins de 40	8	14
40 à moins de 50	25	39
50 à moins de 60	27	66
60 à moins de 70	20	86
70 à moins de 80	12	98
80 à moins de 90	6	104
TOTAL	104	–

2. On utilise le code :

```
> Classes<-c(0,10,20,30,40,50,60,70,80,90)
> Effectifs<-c(0,1,2,3,8,25,27,20,12,6)
> Effcum<-cumsum(Effectifs)
> plot(Classes,Effcum,type='b',main='Courbe des effectifs cumulés croissants',
+ xlab='Classes',ylab='Ecc')
> segments(0,52,55,52,lty='dotted')
> segments(55,52,55,0,lty='dotted')
> text(50,0,Mé)
```

   et on obtient le graphe suivant :



3. On trouve approximativement $m = 55$.
Comme $\frac{n}{2} = \frac{104}{2} = 52$; cela signifie que $m \in [50; 60[$. Avec les encadrements

$$\begin{array}{c|c|c} 39 & \frac{n}{2} = 52 & 66 \\ \hline 50 & m & 60 \end{array}$$

on obtient aisément $m = 50 + \frac{52-39}{66-39}(60 - 50) = 54,81$ à 10^{-2} près par défaut.

4. $\begin{array}{c|c|c} 14 & \frac{n}{4} = 26 & 39 \\ \hline 40 & Q_1 & 50 \end{array} \Rightarrow Q_1 = 40 + \frac{(26-14)}{(39-14)}(50 - 40) = 44,8,$
$\begin{array}{c|c|c} 66 & \frac{3n}{4} = 78 & 86 \\ \hline 60 & Q_3 & 70 \end{array} \Rightarrow Q_3 = 60 + \frac{(78-66)}{(86-66)}(70 - 60) = 66.$

Exercice 23

1. On complète le tableau initial :

Classes (en kg)	Effectifs n_i	x_i	$n_i \nearrow$	$n_i x_i$	$n_i x_i^2$
$[55; 65[$	10	60	10	600	36000
$[65; 75[$	23	70	33	1610	112700
$[75; 77[$	21	76	54	1596	121296
$[77; 79[$	28	78	82	2184	170352
$[79; 81[$	34	80	116	2720	217600
$[81; 83[$	28	82	144	2296	188272
$[83; 85[$	25	84	169	2100	176400
$[85; 95[$	23	90	192	2070	186300
$[95; 105[$	8	100	200	800	80000
TOTAL	200	–	–	15976	1288920

2. Comme $\frac{n}{2} = 100$, la classe modale est $[79; 81[$.

3. $\begin{array}{c|c|c} 33 & \frac{n}{4} = 50 & 54 \\ \hline 75 & Q_1 & 77 \end{array} \Rightarrow Q_1 = 75 + \frac{(50-33)}{(54-33)}(77 - 75) = 76,62$ à 10^{-2} près par excès,
$\begin{array}{c|c|c} 82 & \frac{n}{2} = 100 & 116 \\ \hline 79 & Q_2 = \text{Mé} & 81 \end{array} \Rightarrow \text{Mé} = 79 + \frac{(100-82)}{(116-82)}(81 - 79) = 80,06$ à 10^{-2} près par excès,
$\begin{array}{c|c|c} 144 & \frac{3n}{4} = 150 & 169 \\ \hline 83 & Q_3 = \text{Mé} & 85 \end{array} \Rightarrow Q_3 = 83 + \frac{(150-144)}{(160-144)}(85 - 83) = 83,75.$

Avec le code

```
> Classes<-c(55,65,75,77,79,81,83,85,95,105)
> Centres.Classes<-c(60,70,76,78,80,82,84,90,100)
> Effectifs<-c(0,10,23,21,28,34,28,25,23,8)
> Effcum<-cumsum(Effectifs)
> plot(Classes,Effcum,type='b',main='Courbe des effectifs cumulÃĺs croissants',
```

```
+ xlab='Classes',ylab='Ecc')
> segments(0,50,76.62,50,lty='dotted')
> segments(76.62,50,76.62,0,lty='dotted')
> text(74,0,'Q1')
> segments(0,100,80.06,100,lty='dotted')
> segments(80.06,100,80.06,0,lty='dotted')
> text(78.5,0,'Q2')
> segments(0,150,83.75,150,lty='dotted')
> segments(83.75,150,83.75,0,lty='dotted')
> text(82,0,'Q3')
```

on retrouve les 3 quartiles.

4. $\overline{x} = \frac{15976}{200} = 79,88$.
   $\sigma(X) = \left(\frac{1288920}{200} - (79,88)^2\right)^{\frac{1}{2}} = 7,99$ à 10^{-2} près par excès.

   on retrouve ces valeurs grâce au code R

```
> n<-200
> serie<-rep(Centres.Classes,Effectifs)
> mean(serie)
[1] 79.88
> sd(serie)*sqrt((n-1)/n)
[1] 7.986589
```

Exercice 24

1. On complète le tableau initial :

Classes des salaires	x_i	n_i	$n_i \nearrow$	$n_i \nearrow (\%)$	n_ix_i	$n_ix_i \nearrow$	$n_ix_i \nearrow (\%)$
$[10;12[$	11	40	40	40	440	440	33,85
$[12;14[$	13	30	70	70	390	830	63,85
$[14;16[$	15	20	90	90	300	1130	86,92
$[16;18[$	17	10	100	100	170	1300	100
TOTAL	–	100	–	–	1300	–	–

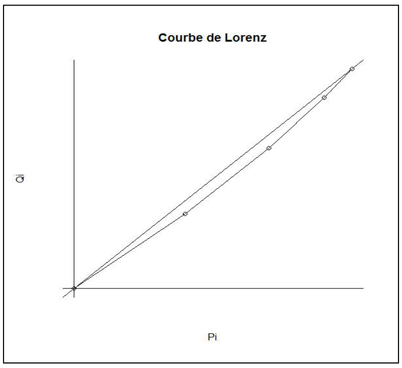
2. On s'intéresse aux colonnes $n_i \nearrow (\%)$ et $n_ix_i(\%)$. Soit p le pourcentage recherché. On a les encadrements suivants :

40	50	70
33,85	p	63,85

   ce qui permet d'affirmer que $p = 33,85 + \frac{(50-40)}{(70-40)}(63,85-33,85) = 43,85$. Le pourcentage de la masse salariale que se partagent les 50% des salariés les mieux payés de cette entreprise est égal à $100 - 43,85 = 56,15\%$.

3. On utilise le code suivant sous R :

```
> Pi<-c(0,40,70,90,100)
> Qi<-c(0,33.85,63.85,86.92,100)
> plot(Pi,Qi,xlim=c(0,100),ylim=c(0,100),main='Courbe de Lorenz',xlab='Pi',ylab='Qi',
+ bty='n',xaxt='n',yaxt='n')
> lines(Pi,Qi)
> abline(h=0)
> abline(v=0)
> abline(0,1)
```



On constate que la répartition des salaires est très égalitaire.

4. Par définition, on a $i = \frac{S}{S_0} = \frac{415,2}{5000} = 0,08$ à 10^{-2} près par défaut. Comme l'indice est proche de 0, la concentration des salaires est faible, ce qui confirme le résultat précédent.

Exercice 25

1. On complète le tableau :

Classes	n_i	$n_i \nearrow$	$P_i\%$	$n_i x_i$	$Q_i\%$
$[1; 1,5[$	52	52	26	65	$17,45$
$[1,5; 2[$	71	123	61,5	$124,25$	$50,8$
$[2; 2,5[$	57	180	90	128,25	$85,23$
$[2,5; 3[$	20	200	100	55	100
TOTAL	200	—	—	372,5	—

2. Comme $\frac{n}{2} = 100$, la salaire médian Mé est dans la classe $[1,5; 2[$. Plus précisément, grâce aux encadrements

1,5	Mé	2
52	100	123

on trouve Mé$= 1,5 + \frac{(100-52)}{(123-52)}(2 - 1,5) = 1,84$ à 10^{-2} près par excès. 50% des salariés de la petite entreprise gagnent moins de 1840 euros.

3. On a $50,8 = \frac{124,25+65}{372,5} \times 100$. $50,8\%$ de la masse salariale est distribuée aux salariés gagnant moins de 2000 euros.

4. On utilise les encadrements suivants :

1,5	Ml	2
17,45	50	50,8

On a alors Ml$= 1,5 + \frac{(50-17,45)}{(50,8-17,45)}(2 - 1,5) = 1,998$ à 10^{-3} près par défaut. 50% de la masse salariale est distribuée aux salariés gagnant moins de 1998 euros.

5. (a) On obtient la courbe de Lorenz.

   (b) Il y a une faible concentration des salaires car la courbe de Lorenz est proche de la première bissectrice. Les salaires sont répartide de manière égalitaire.

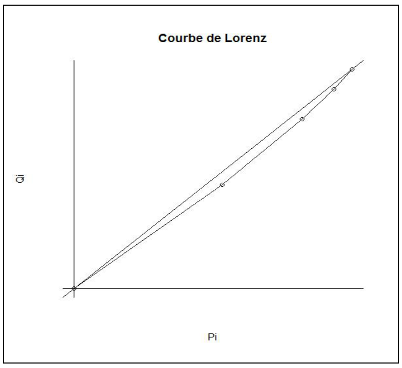
Exercice 26

1.

Salaires en euros	Centres de classes x_i	n_i	$n_i \nearrow$	p_i (%)	$n_i x_i$	$n_i x_i \nearrow$	q_i (%)
$[1000; 1200[$	1100	80	80	53,33	88000	88000	47,21
$[1200; 1400[$	1300	43	123	82	55900	143900	77,20
$[1400; 1600[$	1500	17	140	93.33	25500	169400	90,88
$[1600; 1800[$	1700	10	150	100	17000	186400	100
TOTAL	–	150	–	–	186400	–	–

2. On utilise le code sous `R` :

```
> Classes<-c(1000,1200,1400,1600,1800)
> Centres.Classes<-c(1100,1300,1500,1700)
> Effectifs<-c(80,43,17,10)
> som1<-sum(Effectifs)
> som2<-sum(Effectifs*Centres.Classes)
> Pi<-c(0,cumsum(Effectifs)/som1*100)
> Qi<-c(0,cumsum(Effectifs*Centres.Classes)/som2*100)
> plot(Pi,Qi,xlim=c(0,100),ylim=c(0,100),main='Courbe de Lorenz',xlab='Pi',ylab='Qi',
+ bty='n',xaxt='n',yaxt='n')
> lines(Pi,Qi)
> abline(h=0)
> abline(v=0)
> abline(0,1)
```



La courbe de Lorenz étant très proche de la première bissectrice, la concentration de la masse salariale est faible, la répartition des salaires est égalitaire.