Chapitre 2 : correction des exercices proposés

Exercices pratiques

E.2.1

1906
$\mathcal{X}$: salaire hebdomadaire
$\mathcal{Y}$: nombre hebdomadaire d'heures de travail
$\mathcal{Z}$: salaire horaire

i	x_i	y_i	$z_i = x_i/y_i$
New York	148,95	44	3,39
Londres	54,65	50	1,09
Paris	43,20	54	0,80
Berlin	46,35	53	0,87
Amsterdam	35,00	58	0,60
Rome	20,40	60	0,34
Bruxelles	24,00	60	0,40

1908
$\mathcal{X}$: nombre de téléphones

i	x_i	x_i/total
Europe	2 000 000	27,0 %
Etats-Unis	5 068 800	68,5 %
Canada	130 000	1,8 %
Autres parties du monde	200 000	2,7 %
Total	7 398 800	100 %

1918
x_{it} : prix (en F) de la denrée i à la « date » t ($t = 1$: avril 1914 ; $t = 2$: avril 1916 ; $t = 3$: avril 1917 ; $t = 4$: septembre 1918)

Denrée (i)	Avril 1914 ($t = 1$)	Avril 1916 ($t = 2$)	Avril 1917 ($t = 3$)	Sept. 1918 ($t = 4$)
Bœuf (le kg)	2,50	6,50	11	24
Porc (le kg)	2	7,50	16	25
Pain (les 300g)	0,09	0,14	0,15	0,25
Pommes de terre (le kg)	0,12	?	2	2
Beurre (le kg)	3,50	7,50	18	42
Café (le kg)	3	8	20	85
Lait (le litre)	0,24	0,32	0,80	1,35

E.2.2

Série statistique $\{x_i ; i = 1, \cdots, 10\}$:

3	5	7	0	15	14	4	1	18	6

Série statistique ordonnée $\{x_{(i)} ; i = 1, \cdots, 10\}$:

0	1	3	4	5	6	7	14	15	18

Le moins bon joueur n'a jamais atteint la cible au cours de ses 20 lancers ($x_{(1)} = x_{\min} = 0$).
Le meilleur joueur a atteint 18 fois la cible au cours de ses 20 lancers ($x_{(10)} = x_{\max} = 18$).

E.2.3

Série statistique $\{x_i ; i = 1, \cdots, 20\}$:

2	5	4	2	5	7	0	0	1	3
42	5	3	1	1	0	1	4	4	0

Série statistique ordonnée $\{x_{(i)} ; i = 1, \cdots, 20\}$:

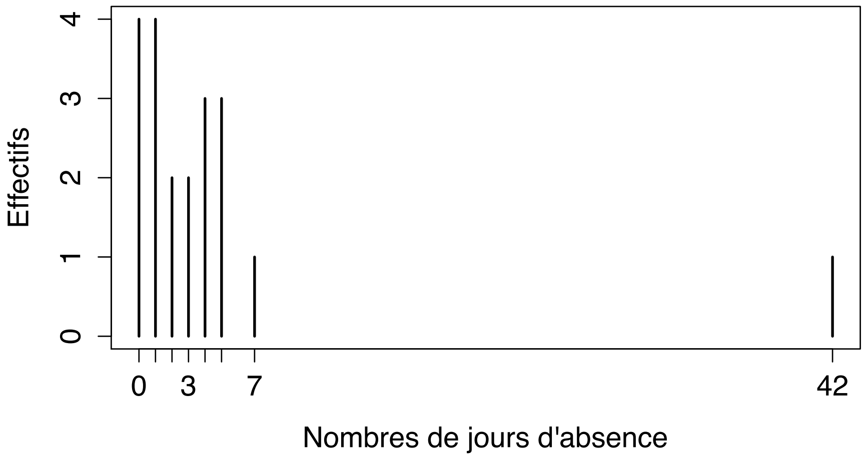
0	0	0	0	1	1	1	1	2	2
3	3	4	4	4	5	5	5	7	42

Distribution observée (D.O.1) :

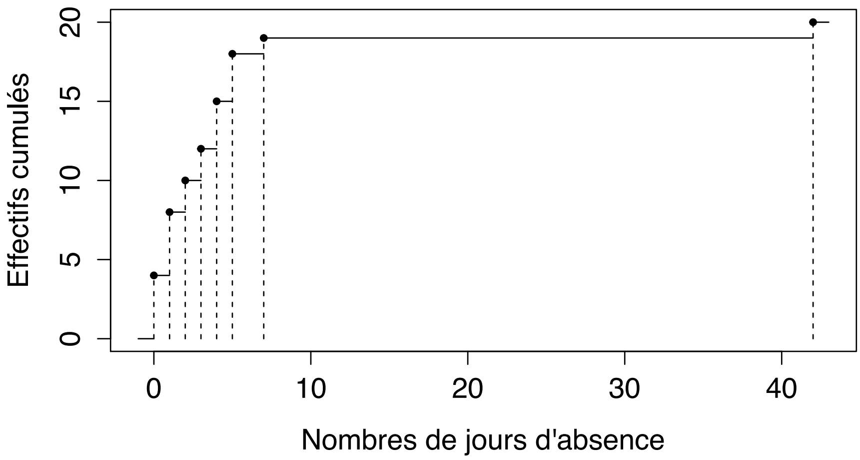
x_j	n_j	f_j	N_j	F_j
0	4	0,20	4	0,2
1	4	0,20	8	0,4
2	2	0,10	10	0,5
3	2	0,10	12	0,6
4	3	0,15	15	0,75
5	3	0,15	18	0,9
7	1	0,05	19	0,95
42	1	0,05	20	1
Total	20	1		

- 4 ouvriers sur les 20 (soit 20% des ouvriers) ne se sont jamais absentés.
- 10 ouvriers sur les 20 (soit 50% des ouvriers) se sont absentés au plus 2 jours.
- 19 ouvriers sur les 20 (soit 95% des ouvriers) se sont absentés au plus 7 jours.
- 1 ouvrier s'est absenté beaucoup plus longtemps que les autres : il comptabilise 42 jours d'absence ; cette valeur de 42 apparaît comme une valeur extrême dans la série observée des nombres de jours d'absence.

Diagramme en bâtons (des effectifs) :



Courbe cumulative (des effectifs) :



E.2.4

Série statistique $\{x_i ; i = 1, \cdots ,20\}$:

59	66	47	37	41	67	84	75	61	72
46	52	56	63	74	62	65	58	77	54

Série statistique ordonnée $\{x_{(i)} ; i = 1, \cdots ,20\}$:

37	41	46	47	52	54	56	58	59	61
62	63	65	66	67	72	74	75	77	84

- Le moins bon résultat est 37 sur 100 ($x_{(1)} = x_{\min} = 37$).
- Le meilleur résultat est 84 sur 100 ($x_{(20)} = x_{\max} = 84$).
- Seulement 4 étudiants sur 20 n'ont pas réussi l'examen (ont obtenu une note inférieure à 50).

E.2.5

Série statistique ordonnée $\{x_{(i)}; i = 1, \cdots, 30\}$:

2	3	3	4	5	6	6	8	8	10	10	11	12	12	13
13	15	17	17	18	19	20	20	21	22	23	24	25	26	27

Distribution observée (D.O.1) :

x_j	n_j	f_j	N_j	F_j
2	1	3,3%	1	3,3%
3	2	6,7%	3	10,0%
4	1	3,3%	4	13,3%
5	1	3,3%	5	16,7%
6	2	6,7%	7	23,3%
8	2	6,7%	9	30,0%
10	2	6,7%	11	36,7%
11	1	3,3%	12	40,0%
12	2	6,7%	14	46,7%
13	2	6,7%	16	53,3%
15	1	3,3%	17	56,7%
17	2	6,7%	19	63,3%
18	1	3,3%	20	66,7%
19	1	3,3%	21	70,0%
20	2	6,7%	23	76,7%
21	1	3,3%	24	80,0%
22	1	3,3%	25	83,3%
23	1	3,3%	26	86,7%
24	1	3,3%	27	90,0%
25	1	3,3%	28	93,3%
26	1	3,3%	29	96,7%
27	1	3,3%	30	100,0%
Total	30	100%		

- Le nombre de jours de vacances passés en dehors du domicile varie entre 2 et 27.
- Pratiquement la moitié des personnes considérées dans l'étude (46,7%) restent absentes de chez elles au plus 12 jours.
- Seulement 7 personnes sur les 30 restent absentes de chez elles plus de 20 jours (c'est-à-dire 21 jours ou plus).

E.2.6

(a) Vrai. Le répondant n° 21 est âgé de 10 ans ; il s’agit du répondant le plus jeune.

(b) Faux. Le répondant n° 5 est âgé de 32 ans ; il n’occupe pas la 5e position dans la série ordonnée des âges. En réalité, $x_{(5)} = x_{15} = 14$.

(c) Vrai. Le répondant n° 29 est âgé de 13 ans et occupe bien la 4e position dans la série ordonnée des âges.

(d) Vrai. L'observation occupant la 10e position dans la série ordonnée des âges vaut bien 21 ; il s'agit de l'âge du répondant n° 36.

(e) Vrai. Le répondant n° 24 occupe bien la 11e position dans la série ordonnée des âges.

(f) Faux. Le répondant occupant la 20e position dans la série ordonnée des âges est le répondant n° 22 ; il est âgé de 40 ans. Donc, $x_{(20)} = 40$.

(g) Vrai. Le répondant n° 25 occupe bien la 25e position dans la série ordonnée des âges.

(h) Vrai. Le répondant n° 20 occupe bien la 21e position dans la série ordonnée des âges et est âgé de 42 ans.

(i) Faux. L'âge occupant la dernière position dans la série ordonnée des âges et l'âge du répondant n° 37. On a donc $x_{(40)} = x_{37} = 95$.

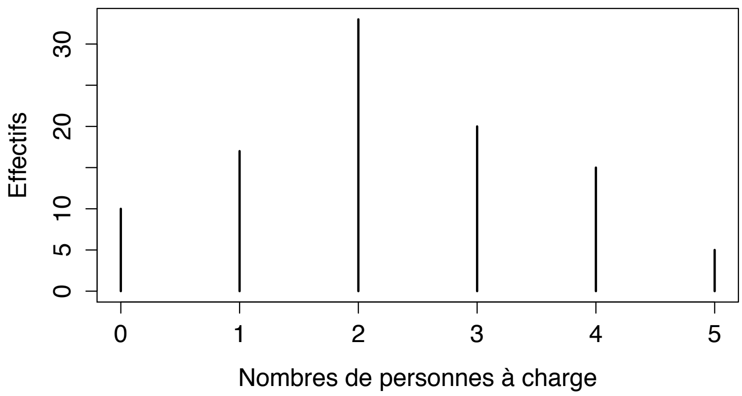
(j) Vrai. L'âge occupant la 39e position dans la série ordonnée des âges vaut 89 ans ; il s'agit de l'âge du répondant n° 40. On a bien ainsi $x_{(39)} = 89 = x_{40}$.

E.2.7

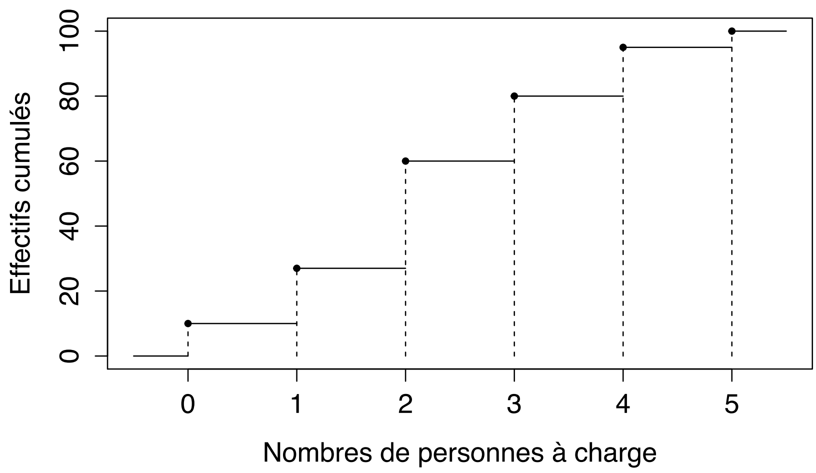
Distribution observée (D.O.1) :

x_j	n_j	f_j	N_j	F_j
0	10	0,10	10	0,10
1	17	0,17	27	0,27
2	33	0,33	60	0,60
3	20	0,20	80	0,80
4	15	0,15	95	0,95
5	5	0,05	100	1
Total	100	1		

Diagramme en bâtons (des effectifs) :



Courbe cumulative (des effectifs) :



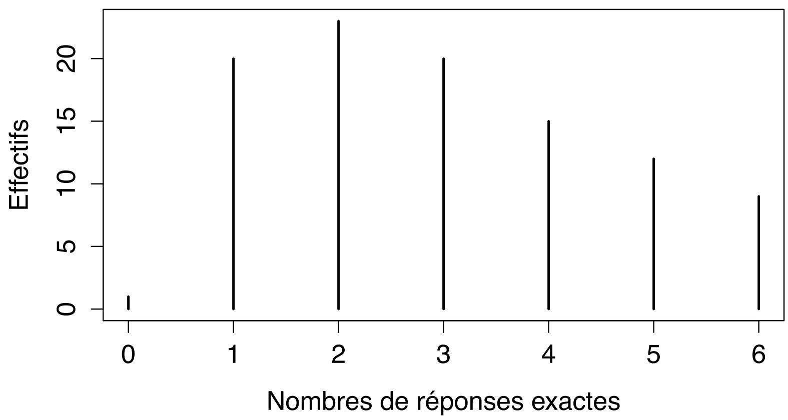
- Le nombre de personnes à charge déclarées varie entre 0 et 5.
- Dans 33% des dossiers, le nombre de personnes à charge déclarées est égal à 2.
- Dans 60% des dossiers, le nombre de personnes à charge déclarées est inférieur ou égal à 2.
- Il n'y a que 20 dossiers sur 100 (soit un cinquième des dossiers) où le nombre de personnes à charge déclarées est supérieur ou égal à 4.

E.2.8

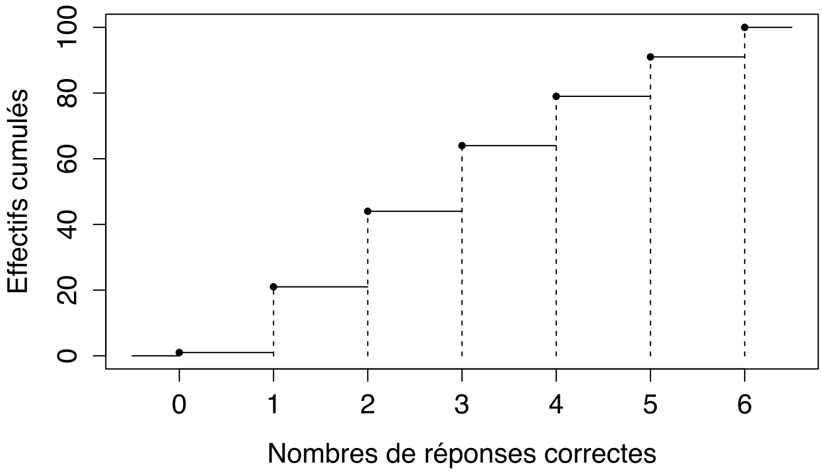
Distribution observée (D.O.1) :

x_j	n_j	f_j	N_j	F_j
0	1	0,01	1	0,01
1	20	0,20	21	0,21
2	23	0,23	44	0,44
3	20	0,20	64	0,64
4	15	0,15	79	0,79
5	12	0,12	91	0,91
6	9	0,09	100	1
Total	100	1		

Diagramme en bâtons (des effectifs) :



Courbe cumulative (des effectifs) :



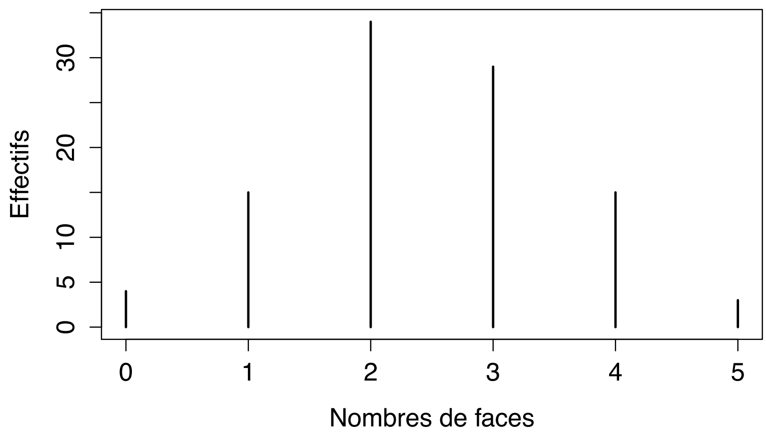
- Une seule copie n'a aucune réponse correcte, alors que 9 copies sur 100 ont donné les réponses correctes aux 6 questions de l'examen.
- 44% des copies n'ont pas obtenu la moitié des points: le nombre de réponses correctes dans ces copies est inférieur ou égal à 2. Le taux d'échec à l'examen est donc élevé!
- 20% des copies ont obtenu juste la moitié des points.
- 36% des copies ont répondu correctement à 4 questions ou plus.

E.2.9

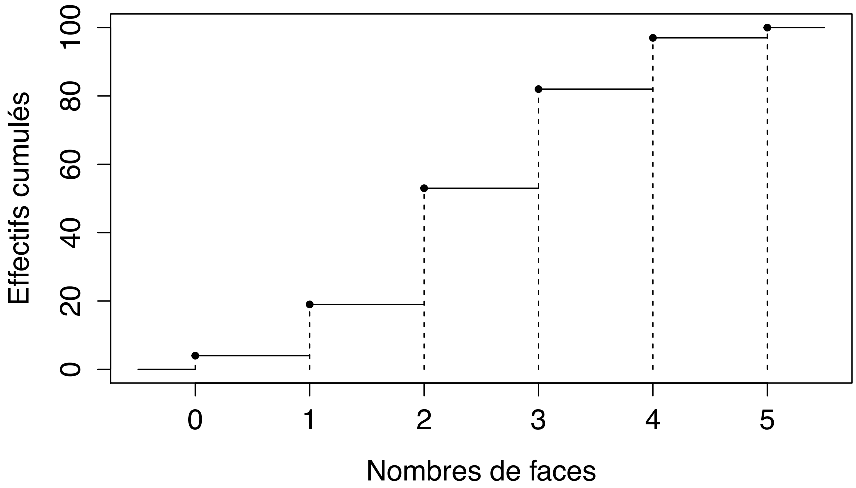
Distribution observée (D.O.1) :

x_j	n_j	f_j	N_j	F_j
0	4	0,04	4	0,04
1	15	0,15	19	0,19
2	34	0,34	53	0,53
3	29	0,29	82	0,82
4	15	0,15	97	0,97
5	3	0,03	100	1
Total	100	1		

Diagramme en bâtons (des effectifs) :



Courbe cumulative (des effectifs) :



- 4 des 100 lancers des 5 pièces de monnaie n'ont donné aucune "face".
- 3 des 100 lancers des 5 pièces de monnaie n'ont en revanche donné que des faces.
- Plus de la moitié des 100 lancers ont conduit à 2 ou 3 faces.
- 82% des lancers ont conduit à au plus 3 faces.

E.2.10

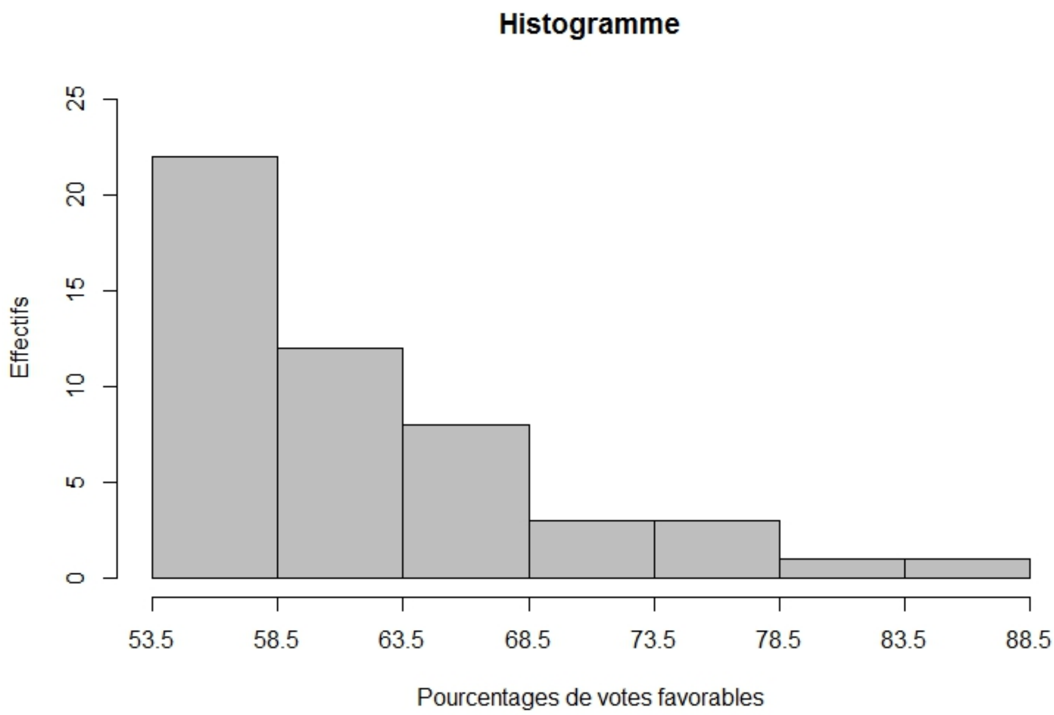
Diagramme en tiges et feuilles (logiciel R) :

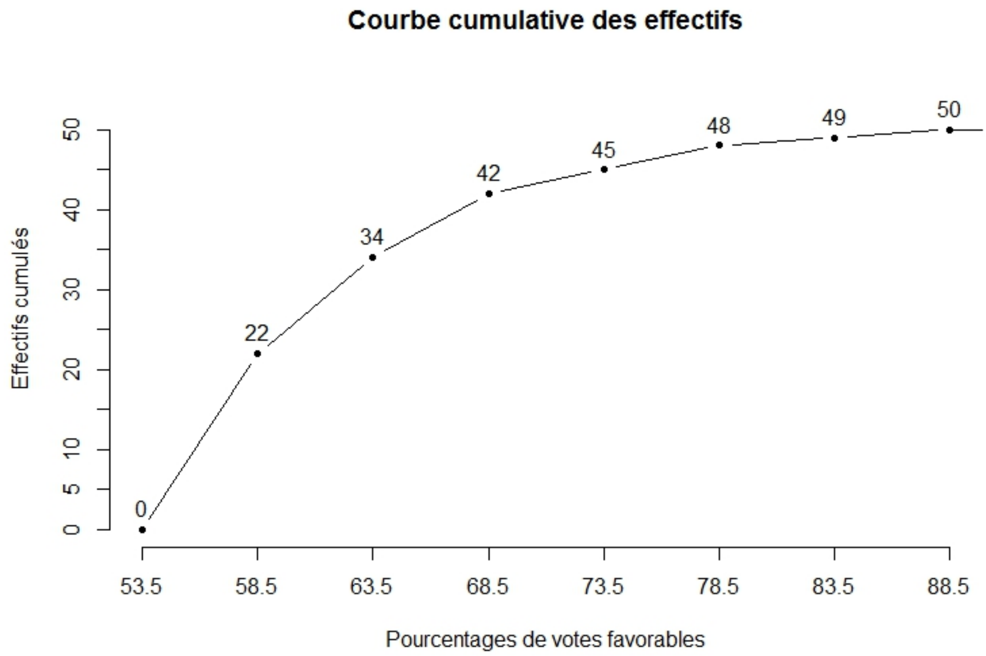
```
5 | 44444
5 | 5555555666777788999
6 | 00112223344
6 | 5567789
7 | 134
7 | 57
8 | 1
8 | 8
```

Distribution groupée (D.G.1) en 7 classes de même longueur :

$\frac{x_{\max}-x_{\min}}{7} = \frac{88-54}{7} = 4{,}86 \Rightarrow$ on va considérer 7 classes de longueur 5

j	C_j	n_j	f_j	N_j	F_j
1	(53,5 - 58,5]	22	0,44	22	0,44
2	(58,5 - 63,5]	12	0,24	34	0,68
3	(63,5 - 68,5]	8	0,16	42	0,84
4	(68,5 - 73,5]	3	0,06	45	0,9
5	(73,5 - 78,5]	3	0,06	48	0,96
6	(78,5 - 83,5]	1	0,02	49	0,98
7	(83,5 - 88,5]	1	0,02	50	1
	Total	50	1		


Histogramme
Effectifs
0
5
10
15
20
25
53.5
58.5
63.5
68.5
73.5
78.5
83.5
88.5
Pourcentages de votes favorables


Courbe cumulative des effectifs
Effectifs cumulés
0
5
10
20
30
40
50
0
22
34
42
45
48
49
50
53.5
58.5
63.5
68.5
73.5
78.5
83.5
88.5
Pourcentages de votes favorables

E.2.11

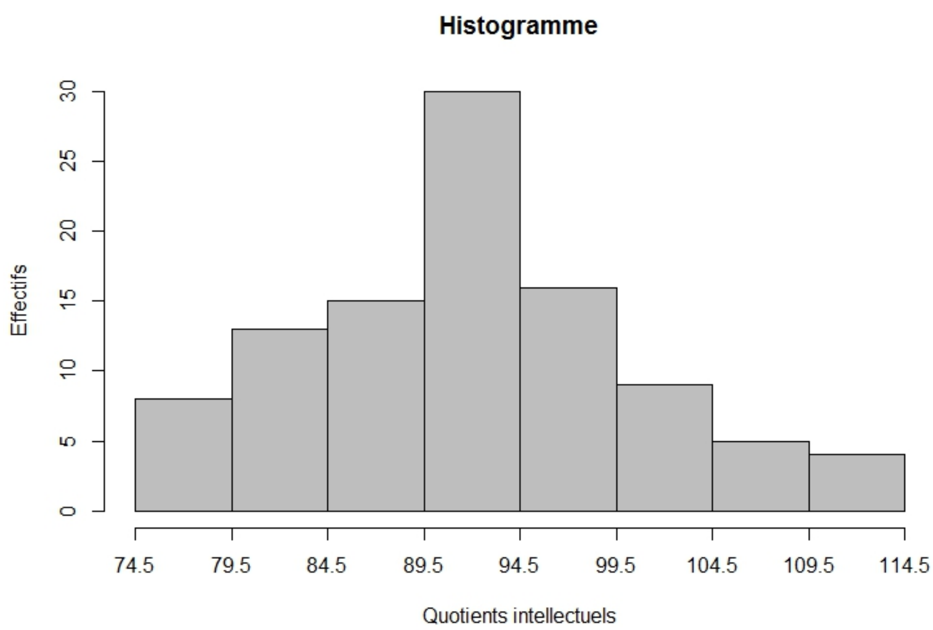
Diagramme en tiges et feuilles (logiciel R) :

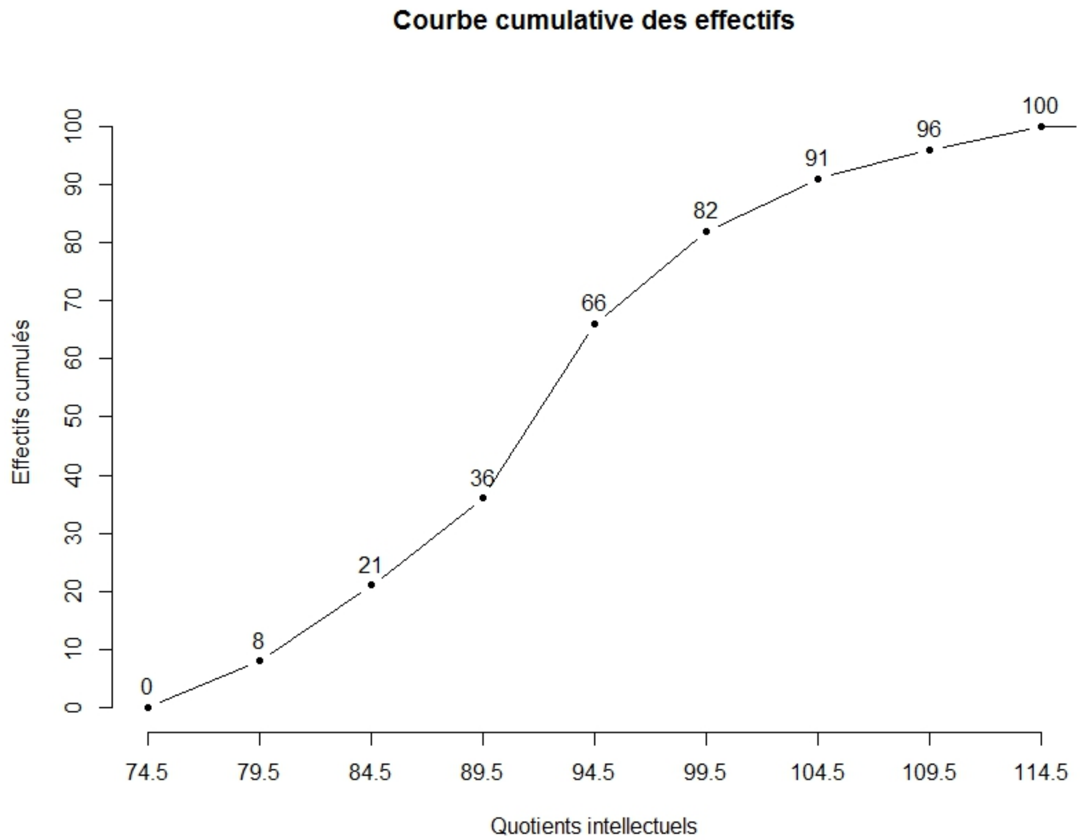
```
 7 | 55666779
 8 | 0002222233444
 8 | 555566677888999
 9 | 000000000011111112222222333344
 9 | 6666667888888899
10 | 000023444
10 | 55567
11 | 0003
```

Distribution groupée (D.G.1) en 8 classes de même longueur :

$\frac{x_{\max}-x_{\min}}{8} = \frac{113-75}{8} = 4{,}75 \Rightarrow$ on va considérer 8 classes de longueur 5

j	C_j	n_j	f_j	N_j	F_j
1	(74,5 - 79,5]	8	0,08	8	0,08
2	(79,5 - 84,5]	13	0,13	21	0,21
3	(84,5 - 89,5]	15	0,15	36	0,36
4	(89,5 - 94,5]	30	0,3	66	0,66
5	(94,5 - 99,5]	16	0,16	82	0,82
6	(99,5 - 104,5]	9	0,09	91	0,91
7	(104,5 - 109,5]	5	0,05	96	0,96
8	(109,5 - 114,5]	4	0,04	100	1
	Total	100	1		



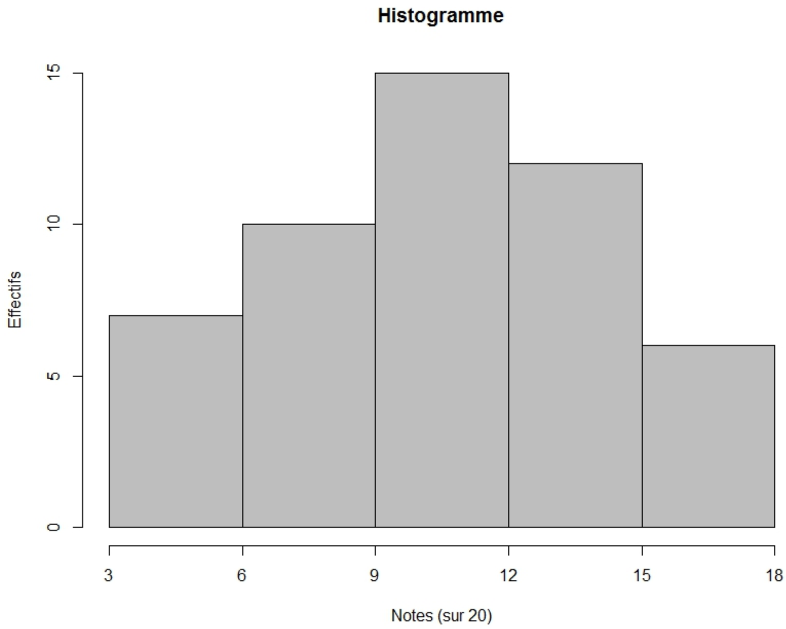


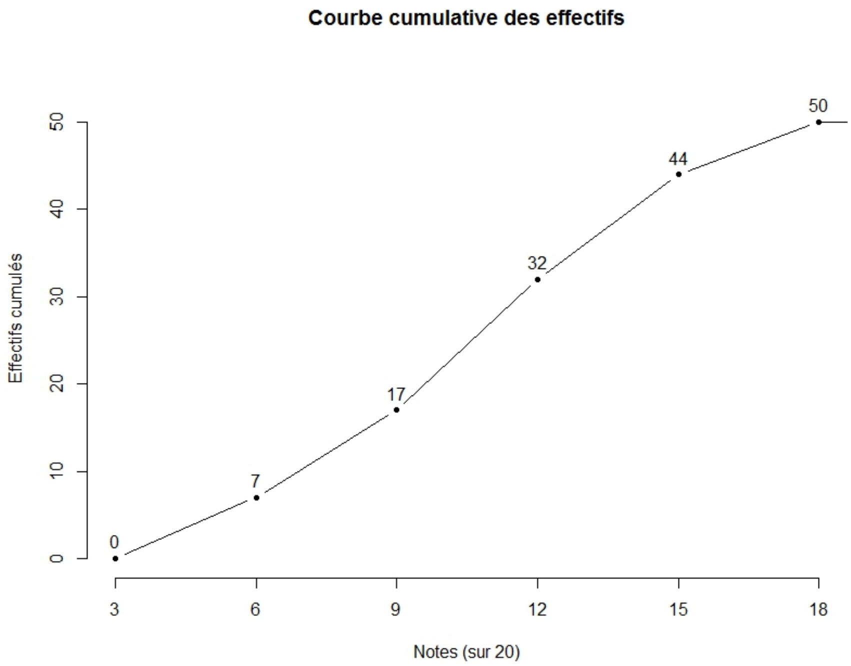
E.2.12

Distribution groupée (D.G.1) en 5 classes de même longueur :

$\frac{x_{\max}-x_{\min}}{5} = \frac{18-3,5}{5} = 2,9 \Rightarrow$ on va considérer 5 classes de longueur 3

j	C_j	n_j	f_j	N_j	F_j
1	(3 - 6]	7	0,14	7	0,14
2	(6 - 9]	10	0,20	17	0,34
3	(9 - 12]	15	0,30	32	0,64
4	(12 - 15]	12	0,24	44	0,88
5	(15 - 18]	6	0,12	50	1
	Total	50	1		


Histogramme
Effectifs
0
5
10
15
3
6
9
12
15
18
Notes (sur 20)

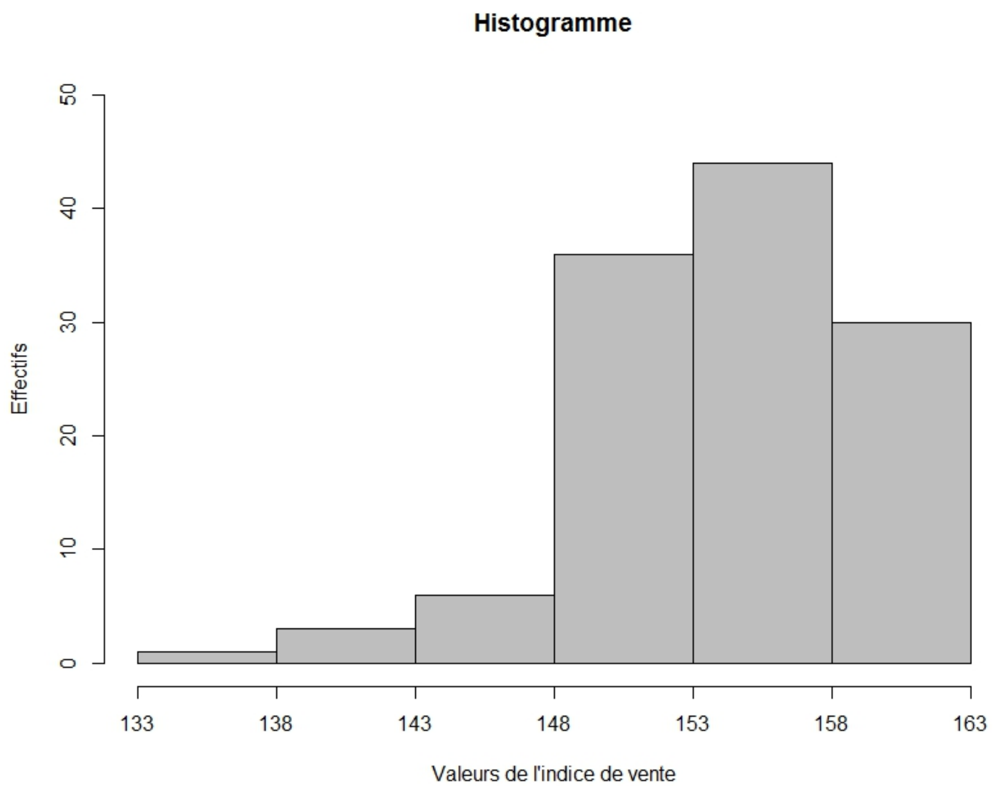

Courbe cumulative des effectifs
Effectifs cumulés
0
10
20
30
40
50
0
7
17
32
44
50
3
6
9
12
15
18
Notes (sur 20)

E.2.13

(i) Distribution groupée (D.G.1) en 6 classes de même longueur :

$\frac{x_{\max}-x_{\min}}{6} = \frac{162-134}{6} = 4{,}67 \Rightarrow$ on va considérer 6 classes de longueur 5

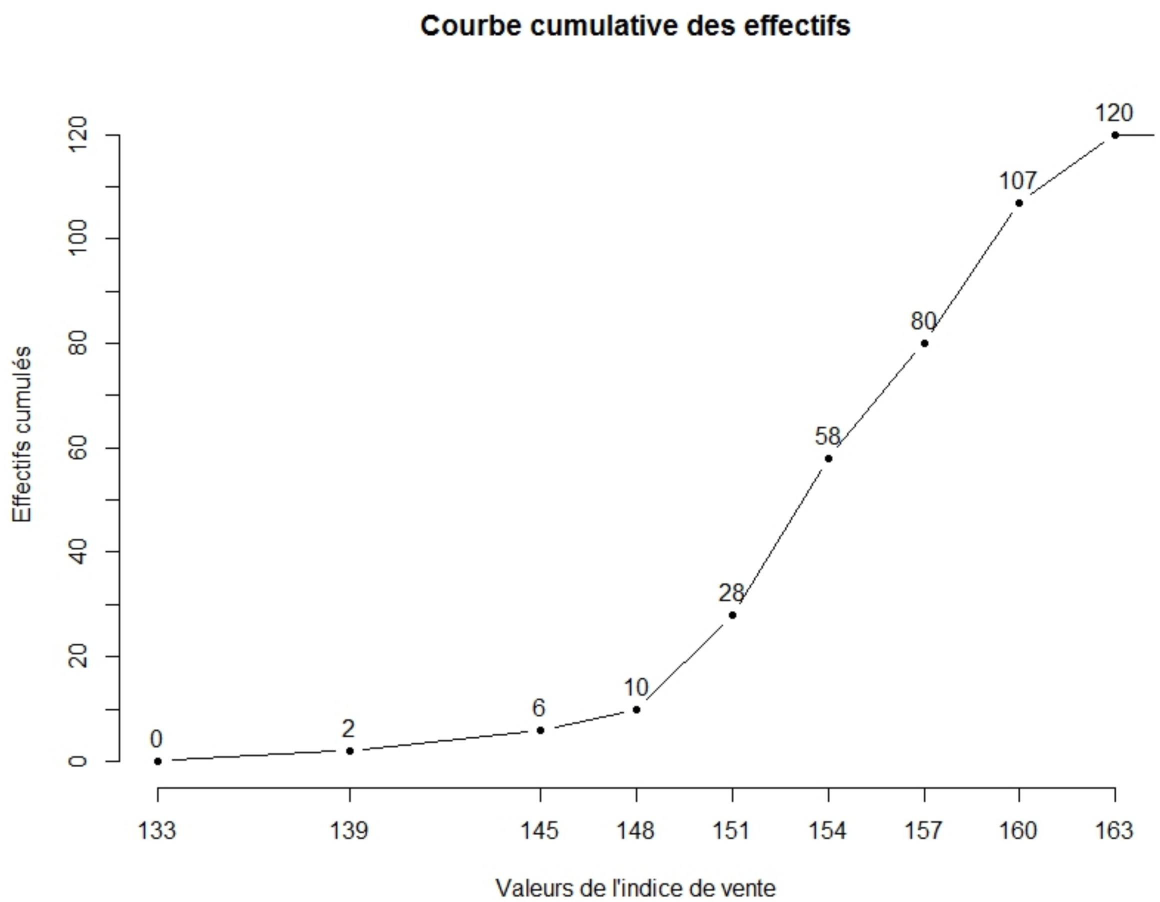
j	C_j	n_j
1	(133 - 138]	1
2	(138 - 143]	3
3	(143 - 148]	6
4	(148 - 153]	36
5	(153 - 158]	44
6	(158 - 163]	30
	Total	120

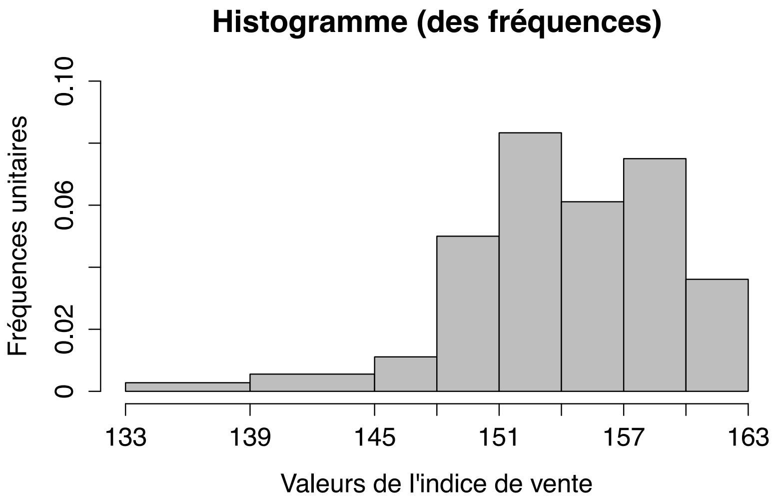


Le découpage en classes que nous venons de considérer ne semble pas très approprié : les trois premières classes ne contiennent que très peu d'observations ; la toute grande majorité des observations se répartissent dans les trois dernières classes. Il apparaît dès lors plus pertinent ici de considérer des classes de longueurs différentes.

(ii) Considérons, par exemple, la distribution groupée (D.G.1) en 8 classes de longueurs différentes suivante :

j	C_j	n_j	f_j	N_j	F_j	h_j	n_j^*	f_j^*
1	[133 - 139]	2	0,017	2	0,017	6	0,33	0,003
2	[139 - 145]	4	0,033	6	0,050	4	1	0,008
3	[145 - 148]	4	0,033	10	0,083	3	1,33	0,011
4	[148 - 151]	18	0,150	28	0,233	3	6	0,050
5	[151 - 154]	30	0,250	58	0,483	3	10	0,083
6	[154 - 157]	22	0,183	80	0,667	3	7,33	0,061
7	[157 - 160]	27	0,225	107	0,892	3	9	0,075
8	[160 - 163]	13	0,108	120	1	3	4,33	0,036
	Total	120	1					



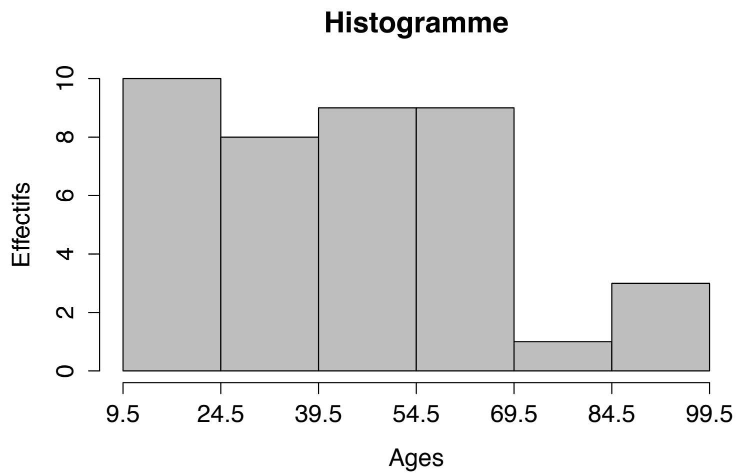


E.2.14

(a) Distribution groupée (D.G.1) en 6 classes de même longueur :

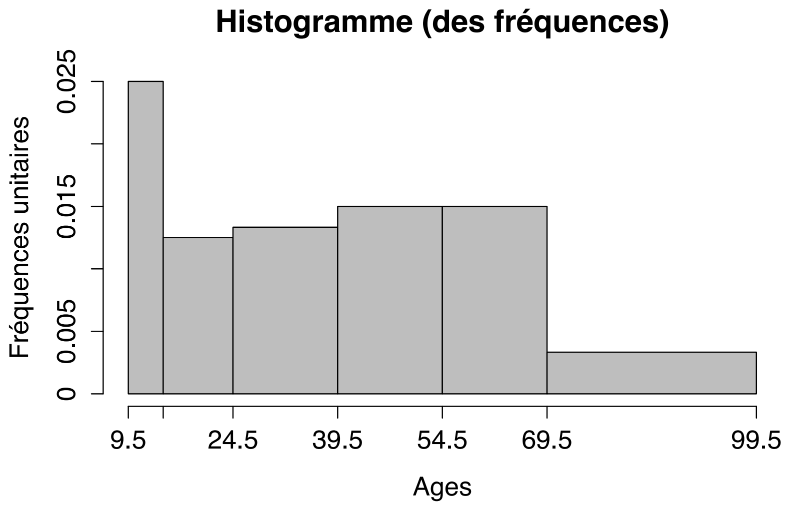
$\frac{x_{\max}-x_{\min}}{6} = \frac{95-10}{6} = 14{,}167 \Rightarrow$ on va considérer 6 classes de longueur 15

j	C_j	n_j	f_j
1	(9,5 - 24,5]	10	0,25
2	(24,5 - 39,5]	8	0,2
3	(39,5 - 54,5]	9	0,225
4	(54,5 - 69,5]	9	0,225
5	(69,5 - 84,5]	1	0,025
6	(84,5 - 99,5]	3	0,075
	Total	40	1



(b) Distribution groupée (D.G.1) en 6 classes de longueurs différentes :

j	C_j	n_j	f_j	h_j	n_j^*	f_j^*
1	(9,5 - 14,5]	5	0,125	5	1	0,025
2	(14,5 - 24,5]	5	0,125	10	0,5	0,0125
3	(24,5 - 39,5]	8	0,2	15	0,53	0,0133
4	(39,5 - 54,5]	9	0,225	15	0,6	0,015
5	(54,5 - 69,5]	9	0,225	15	0,6	0,015
6	(69,5 - 99,5]	4	0,1	30	0,13	0,0033
	Total	40	1			



L'histogramme des effectifs est obtenu en multipliant l'échelle des ordonnées par $n = 40$ de manière à remplacer les *fréquences* unitaires par les *effectifs* unitaires.

E.2.15

Distribution groupée (D.G.1) :

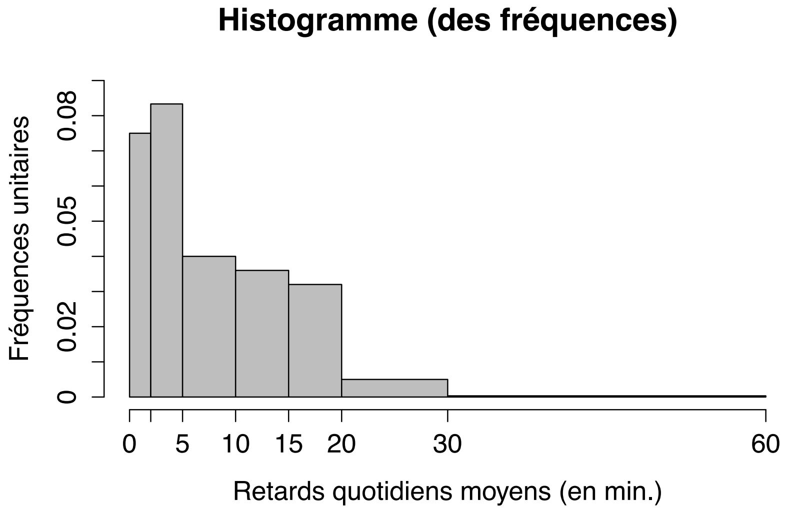
j	C_j	F_j	f_j	n_j	h_j	n_j^*	f_j^*
1	[0 - 2]	15%	15%	30	2	15	0,075
2	(2 - 5]	40%	25%	50	3	16,667	0,0833
3	(5 - 10]	60%	20%	40	5	8	0,04
4	(10 - 15]	78%	18%	36	5	7,2	0,036
5	(15 - 20]	94%	16%	32	5	6,4	0,032
6	(20 - 30]	99%	5%	10	10	1	0,005
7	(30 - 60]	100%	1%	2	30	0,067	0,0003
	Total		100%	200			

(a) Cette fréquence cumulée nous indique que, durant 94% des 200 jours d'observation, le retard quotidien moyen n'a pas dépassé 20 minutes (le retard quotidien moyen est resté inférieur ou égal à 20 minutes durant 94% des jours d'observation).

(b) 60% des retards quotidiens moyens sont inférieurs ou égaux à 10 minutes. La part des retards quotidiens moyens qui dépassent 10 minutes est donc égale à 40%.

(c) 78% des retards quotidiens moyens sont inférieurs ou égaux à 15 minutes et 99% des retards quotidiens moyens sont inférieurs ou égaux à 30 minutes. On en déduit que la part des retards quotidiens moyens qui excèdent 15 minutes mais ne dépassent pas 30 minutes s'élève à 99%-78%, c'est-à-dire à 21% ; cela correspond à 42 jours sur les 200 jours d'observation.

(d)



L'histogramme des effectifs est obtenu en multipliant l'échelle des ordonnées par $n = 200$ de manière à remplacer les *fréquences* unitaires par les *effectifs* unitaires.