

Statistique descriptive  
Année académique 2020–2021  
Carole.Baum@uliege.be

## Chapitre 2 : Organisation et représentation des données

**Exercice 1.** Le tableau 1 décrit, en pourcentages, les résultats d’une enquête menée par le Service Public Fédéral Economie (Statistics Belgium) sur les activités sportives des hommes et femmes en Belgique.

TABLE 1 – Comparaison de l’activité sportive des belges. (*Source : Statistics Belgium*)

	Entraînement intensif (+ 4 h /semaine)	Activités légères (- 4 h /semaine)	Aucune activité sportive
Hommes	27,6%	44%	28,4%
Femmes	8,6%	54,2%	37,2%

- (a) Quelle est la variable étudiée, son type et ses modalités ?
- (b) Construire deux diagrammes en secteurs afin de décrire la répartition des hommes et des femmes en fonction des modalités de la variable.
- (c) En associant à chaque modalité de la variable deux barres côte à côte, une pour les hommes et une pour les femmes, construire un diagramme en barres permettant de comparer sur le même graphique les activités sportives des hommes et des femmes.

**Solution :**

- (a) La variable étudiée est l’activité sportive des Belges. C’est une variable qualitative ordinaire dont les modalités sont : intensive, légère et aucune.
- (b) Voir Figure 1



FIGURE 1 – Diagramme en secteur pour les hommes à gauche et pour les femmes à droite.

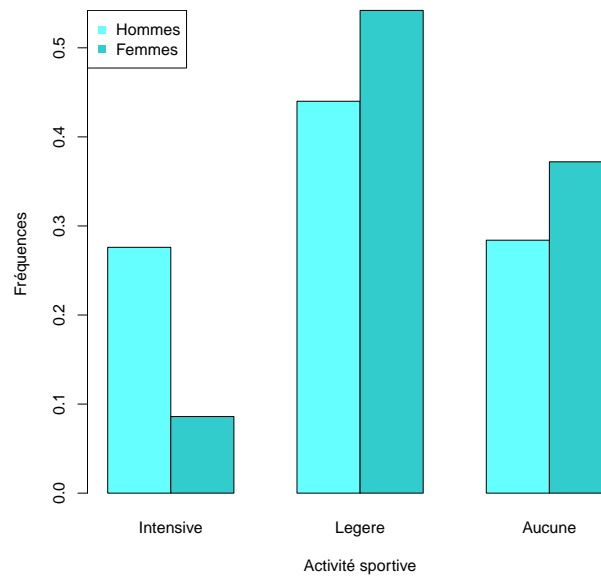


FIGURE 2 – Diagramme en barres.

(c) Voir Figure 2

**Exercice 2.** Le tableau 2 décrit les principales caractéristiques d'activité<sup>1</sup> de la population belge pour l'année 1999.

TABLE 2 – Activités de la population belge. (*Source : Statistics Belgium*)

Modalités	Effectifs (en milliers)
Actif occupé	4007
Chômeur	375
Enfants (de moins de 15 ans)	1805
Non actif de 15 ans à 64 ans	2356
Non actif de 65 ans et plus	1670

- (a) Représenter la distribution des fréquences à l'aide d'un diagramme en secteurs en indiquant les pourcentages sur les secteurs.
- (b) Représenter la série à l'aide d'un diagramme en barres.

**Solution :**

- (a) Voir Figure 3
- (b) Voir Figure 4

1. Selon le Bureau International du Travail, est chômeuse toute personne de 15 ans et plus qui est sans travail, disponible pour travailler et à la recherche d'un emploi.

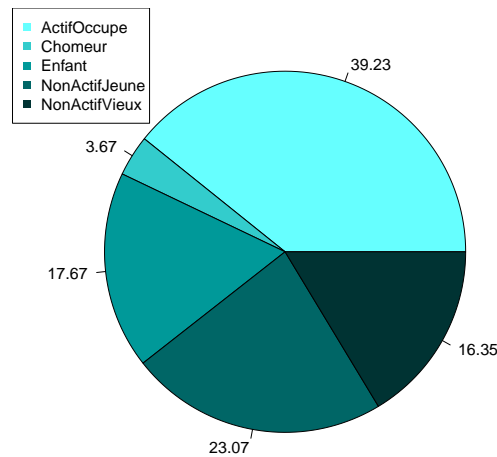


FIGURE 3 – Diagramme en secteur représentant l'activité de la population belge.

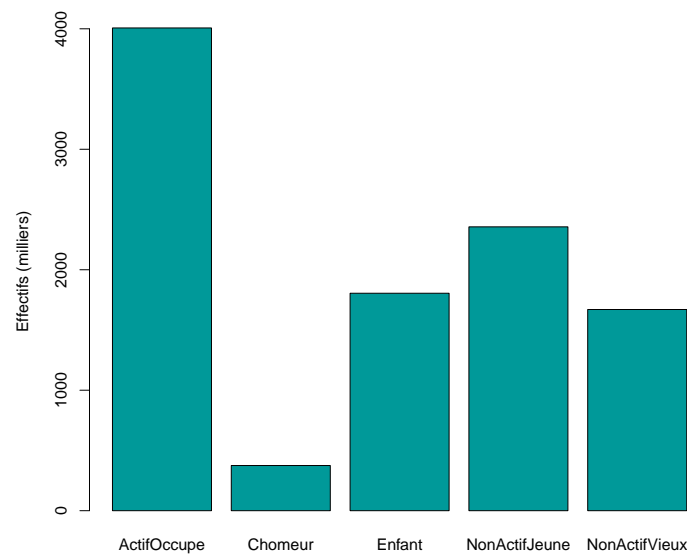


FIGURE 4 – Diagramme en barres représentant l'activité de la population belge.

**Exercice 3.** Lors d'une enquête, on a interrogé 100 travailleurs afin de connaître le nombre de personnes qu'ils avaient à charge. Les données recueillies sont les suivantes :

0	3	1	2	4	3	3	3	4	1	4	1	3	2	2
1	5	2	5	2	4	2	0	3	3	2	4	1	5	2
5	2	5	1	3	2	2	2	4	4	5	1	2	4	2
1	2	2	1	3	5	1	0	3	2	3	1	4	2	4
5	4	1	2	0	2	1	2	2	3	5	2	1	4	2
2	4	0	3	3	2	2	2	0	4	1	2	1	3	1
1	5	3	1	3	2	2	0	2	0					

(a) Établir un tableau statistique de 5 colonnes : Valeurs – Effectifs – Effectifs cumulés – Fréquences – Fréquences

cumulées.

- (b) Représenter la répartition de la population étudiée par un diagramme adéquat basé sur les effectifs.
- (c) Dessiner la courbe cumulative des fréquences cumulées.

**Solution :**

- (a) Tableau statistique :

Valeurs	Effectifs	Effectifs cumulés	Fréquences	Fréquences cumulées
0	8	8	0,08	0,08
1	19	27	0,19	0,27
2	32	59	0,32	0,59
3	17	76	0,17	0,76
4	14	90	0,14	0,9
5	10	100	0,1	1

- (b) Voir Figure 5

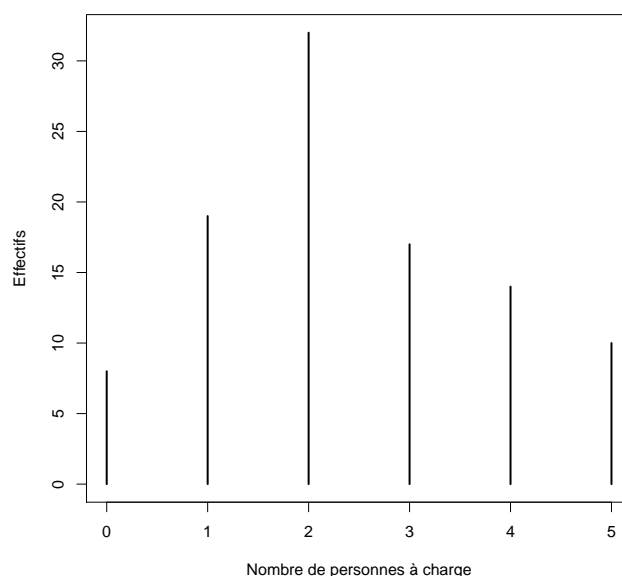


FIGURE 5 – Diagramme en bâtons représentant le nombre de personnes à charge.

- (c) Voir Figure 6

**Exercice 4.** Le tableau 3 donne la répartition des ménages d'un village en fonction du nombre de personnes constituant le ménage.

- (a) Quelle est la variable étudiée ? En déterminer son type et ses modalités. Quelle est la population considérée ?
- (b) Que représente l'effectif total  $n$  de la population étudiée ?
- (c) Combien d'habitants y a-t-il dans la ville considérée ?
- (d) Compléter le tableau des effectifs avec les fréquences, les effectifs et fréquences cumulés.
- (e) A partir du tableau, déterminer le pourcentage de ménages constitués d'au plus 2 personnes et le pourcentage de ménages constitués d'au moins 4 personnes.
- (f) Représenter la distribution des fréquences à l'aide d'un diagramme en bâtons.
- (g) Dessiner la courbe cumulative des fréquences cumulées.

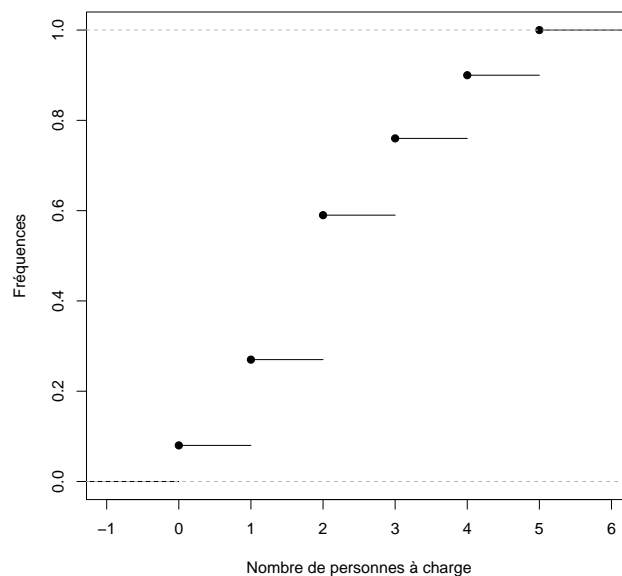


FIGURE 6 – Courbe cumulative des fréquences cumulées représentant le nombre de personnes à charge.

TABLE 3 – Répartition des ménages en fonction du nombre de personnes

Valeurs	Effectifs
1	361
2	453
3	227
4	209
5	83
6	59

**Solution :**

- (a) La variable est le nombre de personnes d'un ménage. C'est une variable quantitative discrète qui prend des valeurs de 1 à 6. La population étudiée est l'ensemble des ménages du village.
- (b) L'effectif total  $n = 361 + 453 + 227 + 209 + 83 + 59 = 1392$  est le nombre de ménages interrogés.
- (c) Il y a  $N = 1 \cdot 361 + 2 \cdot 453 + \dots + 6 \cdot 59 = 3553$  habitants dans la ville considérée.
- (d) Tableau statistique :

Valeurs	Effectifs	Effectifs cumulés	Fréquences	Fréquences cumulées
1	361	361	0,259	0,259
2	453	814	0,325	0,584
3	227	1041	0,163	0,747
4	209	1250	0,15	0,897
5	83	1333	0,06	0,957
6	59	1392	0,043	1

- (e) Pourcentage de ménages constitués d'au plus 2 personnes : 58,4 %  
 Pourcentage de ménages constitués d'au moins 4 personnes :  $(1 - 0.747) \cdot 100 = 25.3\%$
- (f) Voir Figure 7
- (g) Voir Figure 8

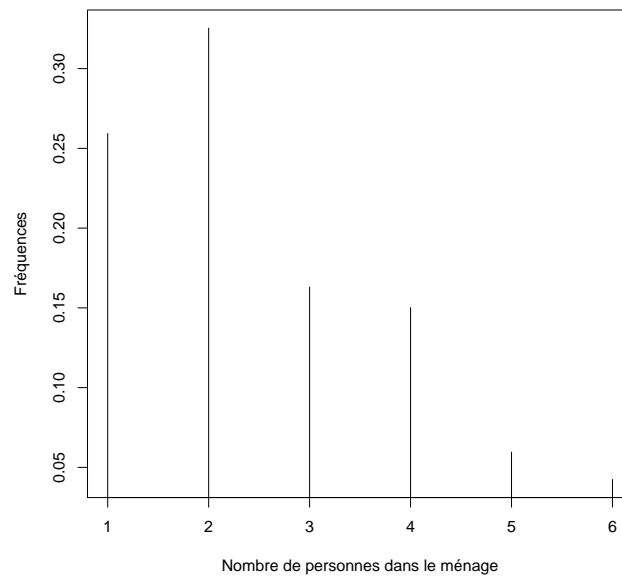


FIGURE 7 – Diagramme en bâtons représentant le nombre de personnes dans le ménage.

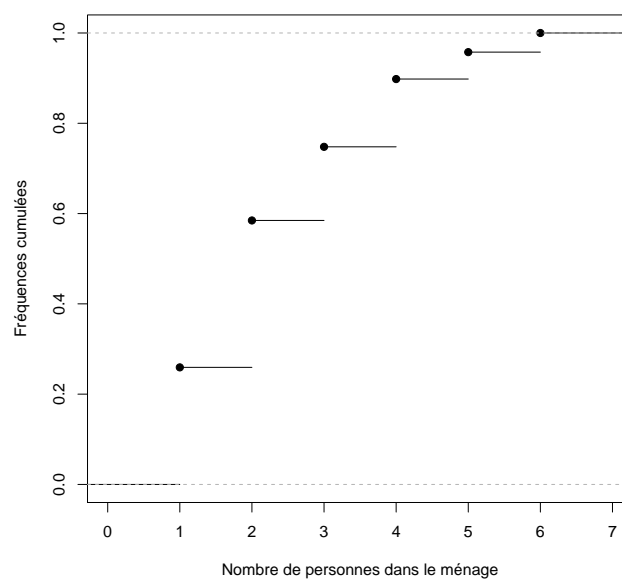


FIGURE 8 – Diagramme en bâtons représentant le nombre de personnes dans le ménage.

**Exercice 5.** Les observations relatives à la variable “Poids” obtenues sur une population d’étudiants masculins de premier bachelier sont transcrites dans le tableau suivant.

68	70	67	75	72	71	67	65	60	60	65	65
77	95	85	70	70	72	66	75	90	65	62	70
52	60	59	65	68	71	97	65	57	75	77	75
85	56	77	67	62	52	67	72	79	60	72	69
58	55	75	75	78	65	95	65	90	72	72	60

- (a) Répartir les 60 données correspondant aux étudiants masculins en les 5 classes d’amplitude constante (égale à 10 kg)

$$\mathcal{C}_1 = [50, 60], \mathcal{C}_2 = ]60, 70], \mathcal{C}_3 = ]70, 80],$$

$$\mathcal{C}_4 = ]80, 90] \text{ et } \mathcal{C}_5 = ]90, 100].$$

Établir pour ces données groupées un tableau statistique de 5 colonnes : Classes – Effectifs – Effectifs cumulés – Fréquences – Fréquences cumulées et représenter la distribution des fréquences par un histogramme et le polygone des fréquences. Le groupement effectué vous semble-t-il adéquat ?

- (b) Améliorer le groupement décrit en (a) en divisant certaines classes et en regroupant d'autres (*suggestion : un tel groupement comprend, par exemple, 6 classes d'amplitudes variables et d'unité d'amplitude égale à 5 kg*). Refaire le tableau statistique, l'histogramme et le polygone des fréquences pour cette nouvelle répartition en classes.
- (c) Construire l'ogive des fréquences cumulées pour les données groupées obtenues en (b). Estimer le nombre d'étudiants dont le poids se situe entre 67.5 kg et 77.5 kg. Comparer l'estimation obtenue avec la vraie valeur. Si les étudiants sont classés par poids croissants, estimer à partir de l'ogive le poids du 30ème étudiant. Comparer cette estimation avec le vrai poids.

### Solution :

- (a) Tableau statistique :

Classes	Effectifs	Effectifs cumulés	Fréquences	Fréquences cumulées
[50, 60]	12	12	0,20	0,20
]60, 70]	22	34	0,37	0,57
]70, 80]	19	53	0,31	0,88
]80, 90]	4	57	0,07	0,95
]90, 100]	3	60	0,05	1
Total	60		1	

L'histogramme et le polygone se trouvent en Figure 9. Comme 68% des données sont dans deux classes et qu'une des classes contient à peine 5%, la répartition choisie n'est pas idéale.

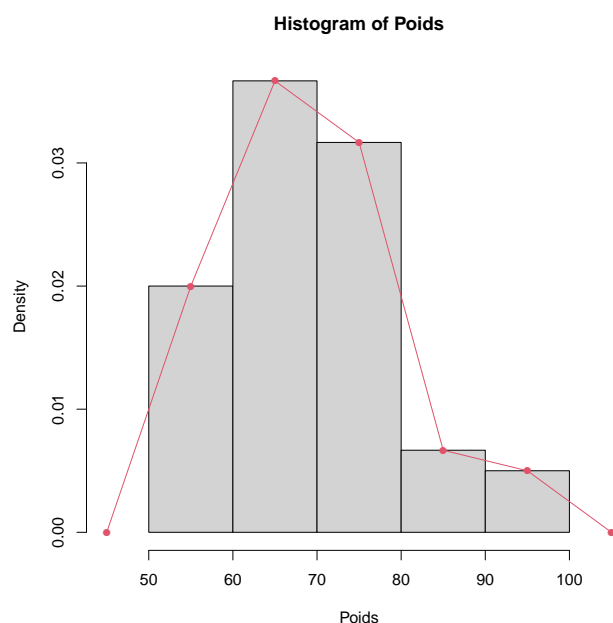


FIGURE 9 – Histogramme des fréquences de la variable Poids

(b) Tableau statistique :

Classes	Eff.	Eff. cumulés	Fréq.	Fréq. cumulées	$a_j$	Fréq. ajustée
[50, 60]	12	12	0,20	0,20	10	0,02
]60, 65]	10	22	0,17	0,37	5	0,034
]65, 70]	12	34	0,20	0,57	5	0,04
]70, 75]	14	48	0,23	0,80	5	0,046
]75, 80]	5	53	0,08	0,88	5	0,016
]80, 100]	7	60	0,12	1	20	0,006

L'histogramme d'aire unitaire correspondant est donné à la Figure 10. Comme les classes ne sont pas de même amplitude, la hauteur des rectangles de l'histogramme doit être basée sur la fréquence ajustée. Celle-ci est calculée en divisant la fréquence de la classe par la largeur de la classe. Les fréquences ajustées sont reprises dans le tableau statistique précédent.

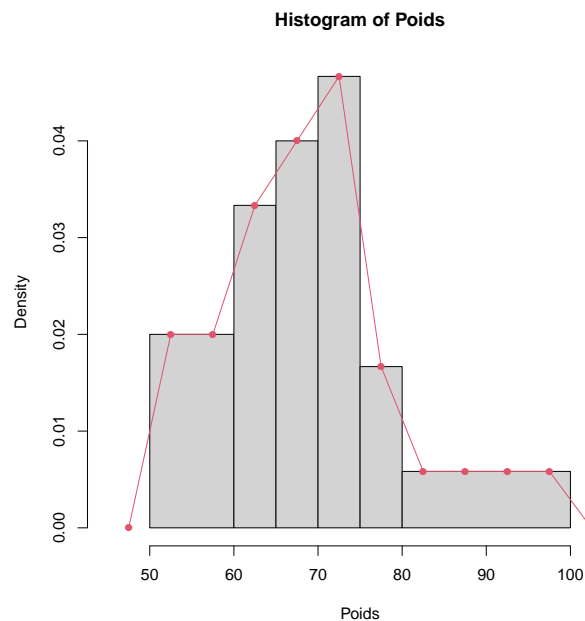


FIGURE 10 – Histogramme des fréquences de la variable Poids

(c) L'ogive des fréquences cumulées est reprise à la Figure 11.

Pour estimer le nombre d'étudiants dont le poids se situe entre 67.5kg et 77.5kg, il faut trouver la fréquence cumulée associée à ces deux poids. Pour trouver ces fréquences cumulées, nous utilisons l'ogive des fréquences cumulées.

Pour commencer, cherchons la fréquence cumulée associée au poids 67.5kg. Dans l'ogive, les points (65; 0.37) et (70; 0.57) sont alignés sur une même droite. Une équation de cette droite nous donne alors

$$y - 0.37 = \frac{0.57 - 0.37}{70 - 65}(x - 65). \quad (1)$$

En remplaçant  $x$  par 67.5 dans cette équation, nous obtenons la fréquence cumulée associée au poids de 67.5kg qui est de

$$F(67.5) = \frac{0.57 - 0.37}{70 - 65}(67.5 - 65) + 0.37 = 0.47.$$

La fréquence cumulée associée au poids de 77.5kg peut être obtenue de la même manière. Mais il est aussi possible de remarquer que l'abscisse 77.5 se situe au milieu du segment reliant les points (75; 0.8) et (80; 0.88). Son ordonnée sera donc  $F(77.5) = \frac{0.8 + 0.88}{2} = 0.84$ .

Le nombre d'étudiants dont le poids se situe entre 67.5kg et 77.5kg est finalement estimé par

$$60 \cdot (F(77.5) - F(67.5)) = 60 \cdot (0.84 - 0.47) = 22.2.$$



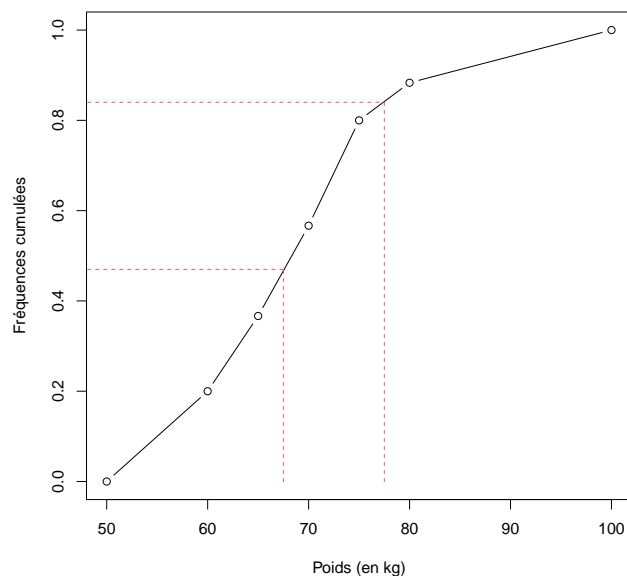


FIGURE 11 – Ogive des fréquences cumulées de la variable Poids

Le vrai nombre d'étudiants dont le poids se situe entre 67.5kg et 77.5kg est de 24 étudiants. Pour trouver cette valeur, il suffit de compter directement dans les données de départ le nombre d'observations comprises entre 67.5 et 77.5.

Le poids du 30ème étudiant est le poids médian. Pour retrouver ce poids dans l'ogive, il faut chercher l'abscisse associée à l'ordonnée  $30/60 = 0.5$ . Le point  $(x; 0.5)$  se situe lui aussi sur le segment de droite donné par l'équation (1). Le poids du 30ème étudiant est donc estimé par

$$x = 65 + \frac{5}{0.2} \cdot 0.13 = 68.25\text{kg}.$$

En réalité, le poids du 30ème étudiant est de 69kg.

**Exercice 6.** Le tableau 4 décrit la répartition des habitants d'une région en fonction du salaire annuel net.

TABLE 4 – Répartition des salaires dans la population active d'une région

Classes (en milliers d'UM)	Effectifs
[0, 10]	181
]10, 15]	116
]15, 20]	89
]20, 25]	39
]25, 30]	28
]30, 35]	60
]35, 50]	91
]50, 60]	58
]60, 100]	65

- Compléter le tableau statistique en ajoutant les fréquences, les effectifs cumulés et les fréquences cumulées des classes.
- Représenter un histogramme d'aire totale unitaire et représenter le polygone correspondant.
- A partir de l'histogramme, estimer le nombre de personnes dont le salaire annuel net est inférieur à 17500 UM et estimer le nombre de personnes dont le salaire annuel net est compris entre 22500 et 37500 UM.

- (d) Représenter l'ogive des fréquences cumulées. Exploiter l'ogive pour recalculer les estimations demandées au point (c).
- (e) Quel salaire annuel net  $s$  vérifie  $F(s) = 0.5$  où  $F$  est l'ogive représentée au point (d)? Que représente intuitivement ce salaire annuel net  $s$ ?

**Solution :**

- (a) Tableau statistique :

Classes	Eff.	Eff. cum.	Fréq.	Fréq. cum.	$a_j$	Fréq. ajustée
[0, 10]	181	181	0,25	0,25	10	0,025
]10, 15]	116	297	0,16	0,41	5	0,032
]15, 20]	89	386	0,12	0,53	5	0,024
]20, 25]	39	425	0,05	0,58	5	0,01
]25, 30]	28	453	0,04	0,62	5	0,008
]30, 35]	60	513	0,08	0,7	5	0,016
]35, 50]	91	604	0,13	0,83	15	0,009
]55, 60]	58	662	0,08	0,91	10	0,008
]60, 100]	65	727	0,09	1	40	0,002
Total	727		1			

- (b) Comme les classes ne sont pas de même amplitude, la hauteur des rectangles de l'histogramme doit être basée sur la fréquence ajustée. Celle-ci est calculée en divisant la fréquence de la classe par la largeur de la classe. Les fréquences ajustées sont reprises dans le tableau statistique précédent. L'histogramme et le polygone sont repris à la Figure 12.

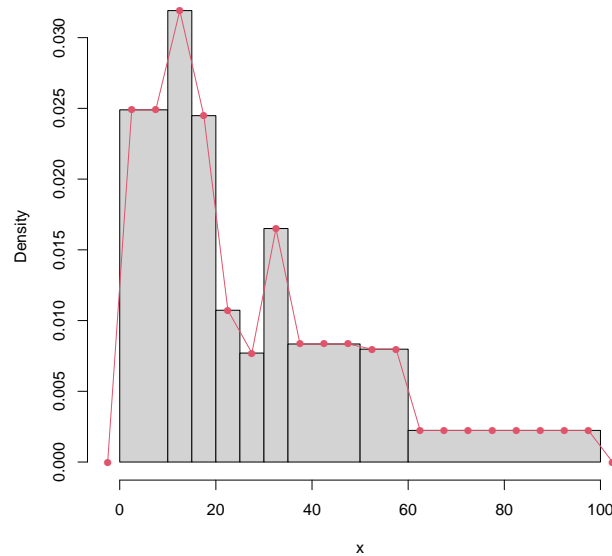


FIGURE 12 – Histogramme de la variable salaire

- (c) Le nombre de personnes dont le salaire annuel net est inférieur à 17500UM est  $181 + 116 + \frac{89}{2} = 342$  personnes.  
Le nombre de personnes dont le salaire annuel net est compris entre 22500 et 37500 UM est  $\frac{39}{2} + 28 + 60 + 91 \cdot \frac{2,5}{15} = 123$  personnes.
- (d) L'ogive des fréquences cumulées est reprise à la Figure 13. Les estimations basées sur l'histogramme et sur l'ogive doivent être identiques car ces deux types d'estimations se basent sur une répartition uniforme au sein des classes.

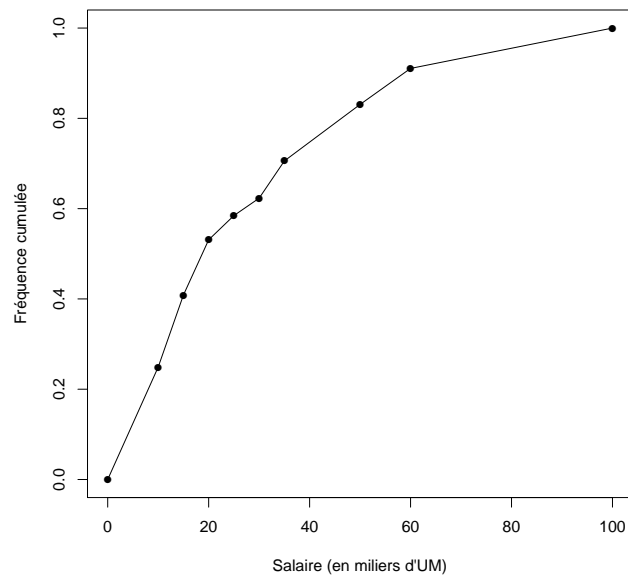


FIGURE 13 – Ogive des fréquences cumulées de la variable salaire

- (e) Le point  $(s; 0.5)$  se trouve sur le segment de droite délimité par les points  $(15; 0.41)$  et  $(20; 0.53)$ . Ce segment a comme équation

$$y - 0.41 = \frac{0.53 - 0.41}{20 - 15}(x - 15)$$

En remplaçant, dans cette équation,  $x$  par  $s$  et  $y$  par  $0.5$ , on trouve

$$s = (y - 0.41) \frac{20 - 15}{0.53 - 0.41} + 15 = 18.75.$$

Le salaire  $s$  tel que  $F(s) = 0.5$  est égal à 18750UM. Il s'agit du salaire annuel médian.