

Compte-rendu TP1 – Statistique Descriptive – 1ère année STID

Attention! Il n'existe pas une manière unique de construire un compte-rendu.

Ce document constitue seulement UN exemple de compte-rendu du T.P.1. De nombreuses autres possibilités sont envisageables!

On considère une population de 534 individus de plus de 16 ans en activité au moment du recensement, en mai 1985.

Le jeu de données contient quatre variables qualitatives (nominales): le sexe (deux modalités), la catégorie professionnelle (six modalités) et les réponses (binaires) aux questions «vivez vous dans le sud?» et «êtes vous membre d'un syndicat?».

Il contient également quatre variables quantitatives dont deux sont discrètes, le nombre d'année d'études et le nombre d'années d'expérience et deux sont continues, le salaire horaire exprimé en dollars et l'âge. Le nombre d'année d'études prend ses valeurs sur l'ensemble $\{0,1,2,\dots,18\}$, le nombre d'années d'expérience sur l'ensemble $\{0,1,2,\dots\}$, le salaire horaire et l'âge sont définis sur \mathbb{R}_+ .

On s'intéresse à décrire la population dans son ensemble au vu de ces huit variables.

La figure 1 représente la répartition de la population selon les différentes catégories professionnelles. Le tableau 1 donne le détail (effectif et fréquence) de la répartition de la population selon cette même variable.

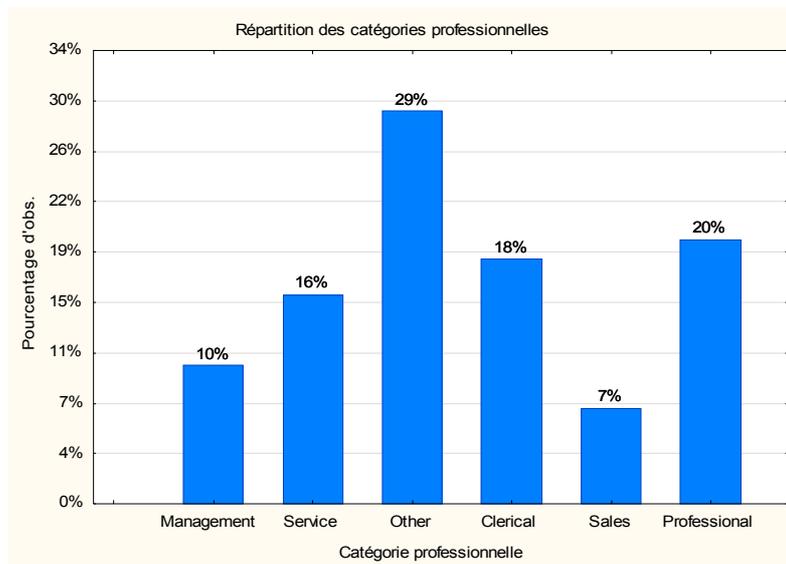


Figure 1 - Répartition des catégories professionnelles

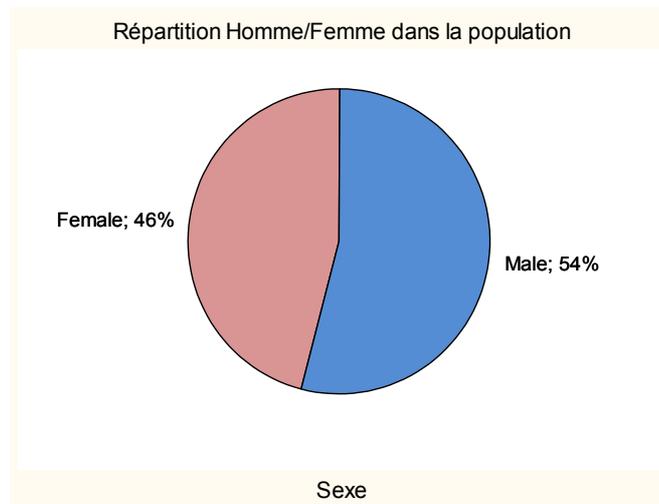
On notera que 105 individus exercent une profession libérale (« Professional »), ce qui représente 20% de la population. Une proportion similaire (18%) est observée pour les employés de bureau (« Clerical »). Les commerciaux (« Sales ») ne représentent que 7% de l'effectif total. Cependant, nous n'avons pas d'information sur l'activité professionnelle d'une part importante (près de 30%) de la population, figurant dans la catégorie « Autre ».

Table 1 - Répartition de la population selon la catégorie professionnelle en effectif et en fréquence

	Effectif	Fréquence
Management	55	10,30
Clerical	97	18,16
Professional	105	19,66
Sales	38	7,12
Service	83	15,54
Other	156	29,21

La population se compose d'un peu plus d'hommes que de femmes (54% contre 46%). Cette observation est visible grâce au diagramme circulaire correspondant à la figure 2.

Figure 2 - Répartition H/F dans la population totale



La figure 3 nous montre que la majorité de la population ne vit pas dans le sud des États-Unis(71%).

Figure 3 - Répartition Nord/Sud dans la population totale



Enfin, la majorité de la population n'est pas membre d'un syndicat (82%).

Table 2- Effectif et fréquence de l'appartenance à un syndicat dans la population totale

	Effectif	%age
No	438	82,02
Yes	96	17,98

Concernant le nombre d'années d'étude, la majorité de la population (environ 40%) a effectué 12 ans d'études. De plus, moins de 20% de la population a étudié moins de 12 ans. Le minimum est de deux années d'études, mais cela ne concerne qu'un seul individu. Près de 6% des individus a effectué 18 ans d'études.

Figure 4 – Effectifs et fréquences cumulées du nombre d'années d'étude dans la population totale.

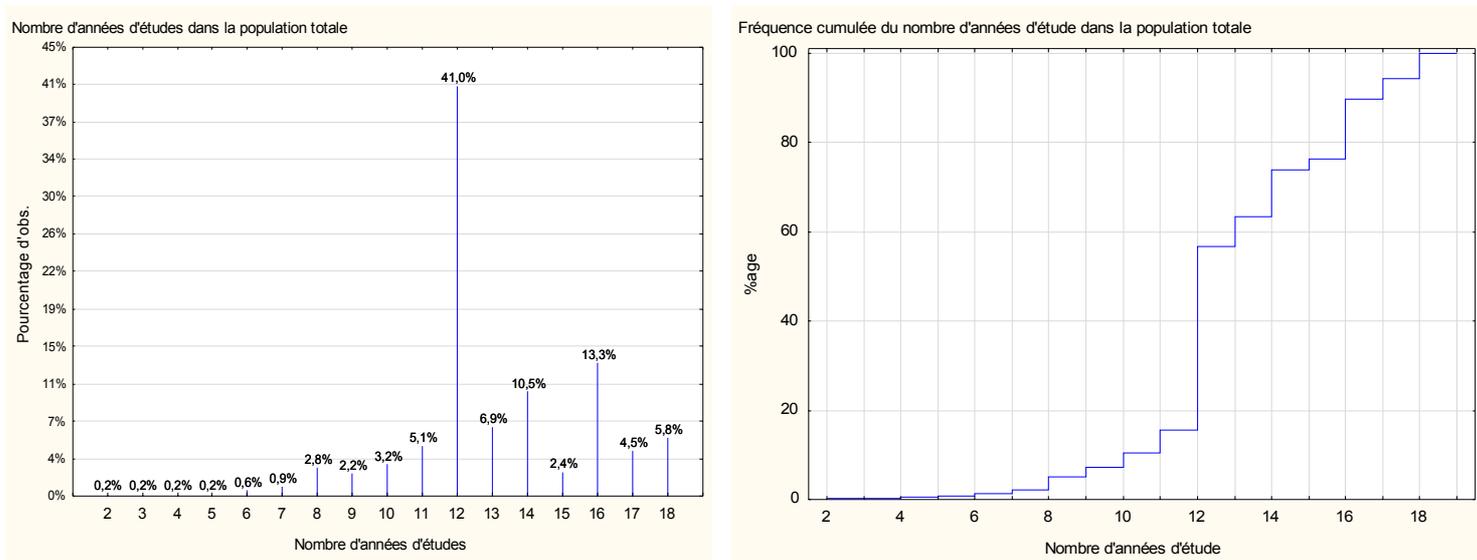


Table 3 - Nombre d'années d'étude dans la population totale

	Effectif	%age	%age - Cumulé
2	1	0.19	0.19
3	1	0.19	0.37
4	1	0.19	0.56
5	1	0.19	0.75
6	3	0.56	1.31
7	5	0.94	2.25
8	15	2.81	5.06
9	12	2.25	7.30
10	17	3.18	10.49
11	27	5.06	15.54
12	219	41.01	56.55
13	37	6.93	63.48
14	56	10.49	73.97
15	13	2.43	76.40
16	71	13.30	89.70
17	24	4.49	94.19
18	31	5.81	100.00

L'âge des individus de cette population en activités lors du recensement varie entre 18 et 68 ans. Près de 60% des individus avaient moins de 38 ans lors de l'étude, et seulement 6% environ avaient plus de 58 ans.

Table 4 - Age des individus (en années) au moment du recensement.

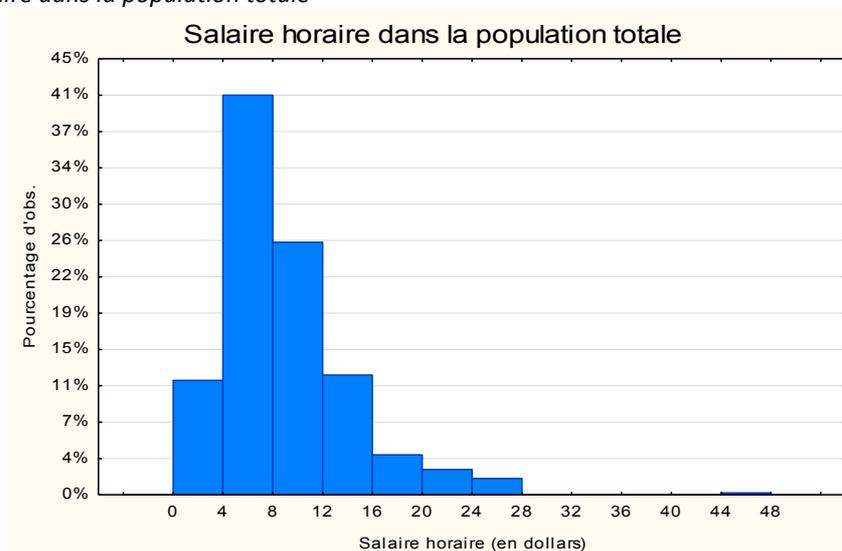
	Effectif	%age	%age - Cumulé
[18,23[53	9,93	9,93
[23,28[77	14,42	24,34
[28,33[93	17,42	41,76
[33,38[97	18,16	59,93
[38,43[60	11,24	71,16
[43,48[46	8,61	79,78
[48,53[33	6,18	85,96
[53,58[42	7,87	93,82
[58,63[22	4,12	97,94
[63,68[11	2,06	100,00

En terme de salaire, environ 90% de la population gagne moins de 16 dollars par heure. Environ la moitié de la population gagne moins de 8 dollars par heure tandis que la majorité des individus (41%) gagnent entre 4 et 8 dollars par heure.

Table 5 - Salaires (dollars par heure) dans la population totale

	Effectif	Pourcentage	Pourcentage Cumulé
]0;4]	63	11,80	11,80
]4;8]	220	41,20	53,00
]8;12]	139	26,03	79,03
]12;16]	66	12,36	91,39
]16;20]	22	4,12	95,51
]20;24]	14	2,62	98,13
]24;28]	9	1,69	99,81
>28	1	0,19	100,00

Figure 6 - Salaire horaire dans la population totale



La représentation de la répartition cumulée du salaire nous donne de nouvelles informations. Celle-ci peut se faire à partir des données brutes, telles qu'elles apparaissent dans la feuille de données de départ, ou à partir des données regroupées en classes (en utilisant alors la Table 5 ci-dessus). La première méthode donne un résultat plus précis, sans interpolation et est bien sûr à privilégier quand on dispose des données brutes, comme dans notre cas.

Cependant, les grandes tendances ressortent sur les deux courbes: près de 80% des individus ont un salaire inférieur à 12 dollars par heure, et près de 50% des individus ont même un salaire horaire ne dépassant pas 10 dollars par heure.

Figure 7 - Fréquence cumulée du salaire horaire dans la population totale (variable groupée)

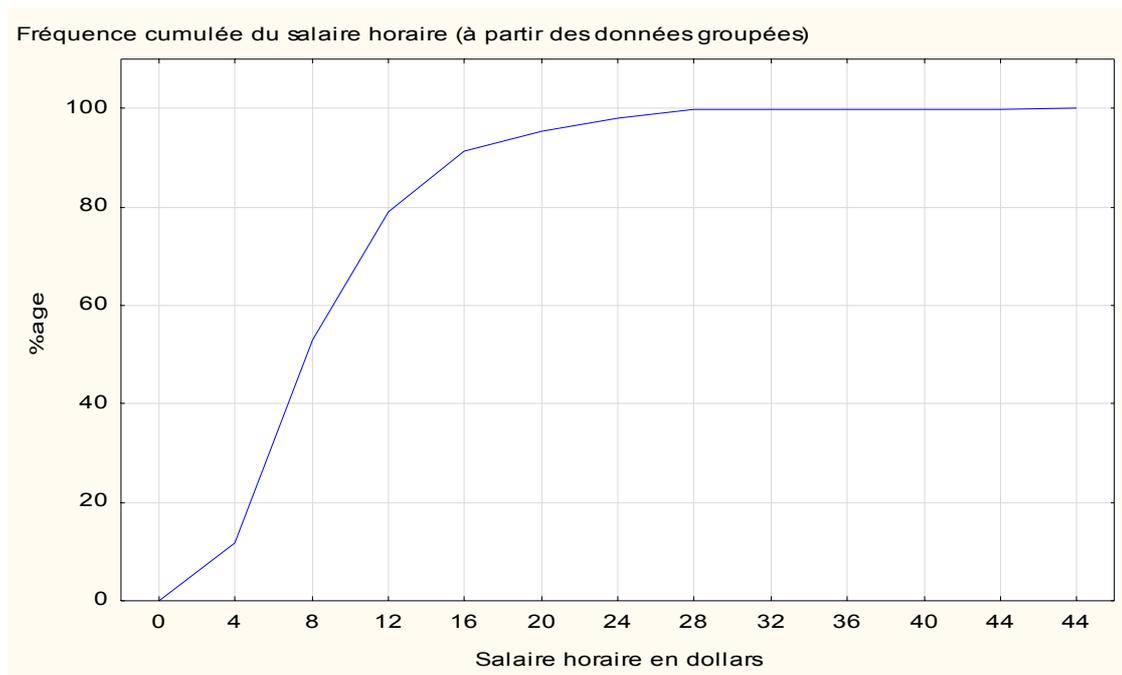
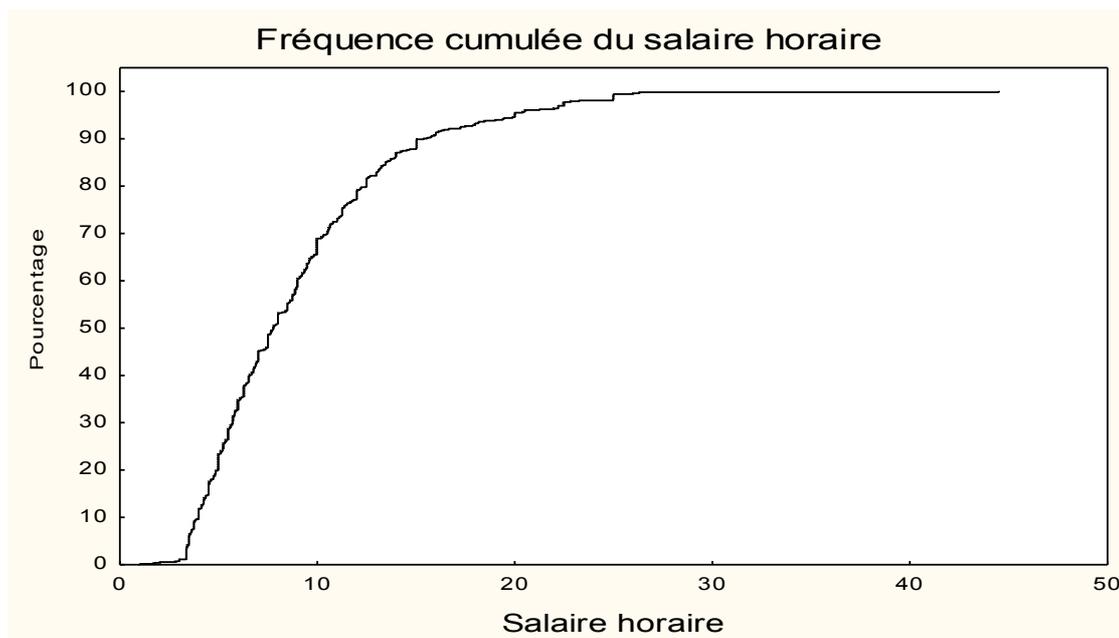


Figure 8 - Fréquence cumulée du salaire horaire dans la population totale (données brutes)



En conclusion, nous avons pu décrire cette population, tant au niveau de la répartition homme/femme, de la localisation géographique du domicile, qu'au niveau professionnel: nombre d'années d'étude, salaire, appartenance à un syndicat.

On a ainsi pu observé que la population étudiée comporte une part homme/femme assez équilibré, des individus vivant très majoritairement dans le Nord des États-Unis et n'appartenant pas à un syndicat. L'information concernant la catégorie professionnelle est plus difficilement exploitable compte tenu de la part non négligeable d'individus pour lesquels elle n'est pas renseignée. On notera quand même une part plus importante des professions libérales et employé de bureau. La majorité des individus a effectué 12 années d'études. Concernant le salaire horaire, seule une minorité des individus ont un salaire supérieur à 20 dollars de l'heure (environ 5% de la population).

Il serait intéressant de rechercher s'il existe des disparités au sein de chaque catégorie professionnelle (en terme de Sexe ou d'âge par exemple), où si les tendances dégagées pour la population dans son ensemble se retrouvent dans chacune d'entre elles, en particulier pour les catégories dont le poids est important par rapport à l'ensemble de la population (professions libérales, employés de bureau par exemple). D'autres pistes d'étude peuvent également être intéressante comme l'étude des variables ayant un impact sur le salaire horaire des individus.