

Objectifs

1-Mettre en ordre les données d'un caractère qualitatif et quantitatif

2-Déterminer et apprécier l'intérêt des fréquences cumulées « moins de »

3-Construire les tableaux d'effectifs pour les données d'une variable

Plan :

1-Définitions

2-Mise en ordre des données

2.1. D'un caractère qualitatif

2.2 D'un caractère quantitatif

2.3 Données manquantes

3-Présentation tabulaire de l'information

Conclusion

Références bibliographiques

1.1. EFFECTIF :

L'effectif ou fréquence absolue est le nombre d'individus appartenant à une modalité donnée.

Prenons l'exemple simple de la distribution de 50 malades selon le sexe. Parmi ces 50 malades, 15 sont de sexe masculin et 35 de sexe féminin. Les effectifs correspondant à chacune des deux modalités sont 15 et 35.

1.2. FREQUENCE RELATIVE :

Une fréquence relative est le rapport entre l'effectif d'une modalité de la variable étudiée et l'effectif total de la population ou plus généralement de l'échantillon sur lequel cette variable est mesurée.

Le numérateur fait obligatoirement partie du dénominateur.

La fréquence relative s'exprime généralement en pourcentage (%).

Pour l'exemple précédent, la fréquence relative au sexe masculin est :

$15/50 = 0,30 = 30\%$ et la fréquence relative au sexe féminin est $35/50 = 0,70 = 70\%$.

1.3. SERIE STATISTIQUE :

Une série statistique est l'ensemble des valeurs prises par une variable qualitative ou quantitative, comme, par exemple, les valeurs du poids d'un groupe d'étudiants ou des durées de séjour d'un groupe de malades hospitalisés.

Exemple de base de données statistiques

No.matric	Etatdos	Sexe	DateN	Age	Adrdetail	Adrcode	DateDiagnos	Localisation	Morphologie	Comportem	Grade	Base
14980	1	2	01/07/1935	80	ANNABA	239999	01/07/2015	189	8140	3	9	7
14998	1	2	06/08/1977	38	ELEULMA	230403	01/07/2015	569	8000	1	9	2
13094	1	1	15/10/1958	56	CT90LOGT ANI	230101	01/01/2015	649	8120	3	3	7
13478	1	2	01/07/1984	31	ANNABA	239999	01/01/2015	169	9591	3		7
13520	1	2	01/07/1959	56	ANNABA	239999	03/01/2015	189	8140	3	9	7
13303	1	1	01/01/1934	81	ANNABA	239999	04/01/2015	619	8140	3	1	7
13304	1	1	01/01/1934	81	ANNABA	239999	04/01/2015	619	8140	3	2	7
13521	1	1	01/01/1992	23	ANNABA	239999	04/01/2015	779	9652	3		4
13522	1	2	06/12/1946	68	ANNABA	239999	04/01/2015	539	8070	3	2	7
13523	1	1	01/01/1961	54	ANNABA	239999	04/01/2015	239	8140	3	1	7
12953	1	2	01/07/1960	55	CTE 5 JUILLET E	230501	05/01/2015	509	8500	3	9	7
13524	1	1	01/01/1980	35	ANNABA	239999	05/01/2015	189	8490	3	9	7
15006	1	1	10/12/1946	68	JARDIN ENFAN	230302	05/01/2015	209	8144	3	2	7
13526	1	1	01/01/1939	76	ANNABA	239999	06/01/2015	349	8140	3	9	7
14459	1	2	00/00/0000	99	ANNABA	239999	06/01/2015	251	8000	3		0
15000	1	2	18/04/1967	47	EL HADJAR	230301	06/01/2015	509	8500	3	3	7
15067	1	1	07/11/1946	68	ANNABA	239999	06/01/2015	349	8071	3	9	7
15829	1	2	17/03/1958	56	504Lg BOUKHA	230201	06/01/2015	509	8500	3	3	7
13305	1	2	01/01/1968	47	EL HADJAR	230301	07/01/2015	739	8330	3	9	7
13306	1	2	01/01/1943	72	ANNABA	239999	07/01/2015	441	8090	3	9	7
13421	1	1	05/02/1950	64	ANNABA	239999	07/01/2015	341	8263	3	9	7
13527	1	2	01/01/1962	53	ANNABA	239999	07/01/2015	509	8500	3	2	7
13528	1	2	01/01/1962	53	ANNABA	239999	07/01/2015	509	8500	3	2	7

2. MISE EN ORDRE DES DONNEES :

La mise en ordre des données est l'étape qui suit immédiatement celle du recueil des valeurs de la (ou des) variable (s) étudiée(s). Elle consiste à dresser un tableau qui fait correspondre aux valeurs ou qualités de la variable prise en considération, le nombre d'individus présentant effectivement ces valeurs ou ces qualités. Le tableau en question s'appelle tableau d'effectifs ou distribution de fréquences.

Il s'agit donc d'effectuer un dépouillement qui va tenir compte de la nature qualitative ou quantitative de la variable étudiée.

2.1- MISE EN ORDRE DES DONNEES D'UN CARACTERE QUALITATIF :

Les modalités d'un caractère qualitatif doivent épuiser toutes les possibilités et ne pas empiéter les unes sur les autres. On dit en d'autres termes qu'elles sont collectivement exhaustives et mutuellement exclusives.

Supposons qu'on ait dressé le tableau d'effectifs de personnes étudiées selon le groupe sanguin dans le système ABO. Les modalités O, A, B, AB, satisfont aux critères ci-dessus énoncés.

Dans le cas où certaines modalités ont une fréquence rare, dans le cas précis d'une étude donnée, on pourra les regrouper sous une rubrique unique qui sera appelée « autre ».

Considérons l'exemple simple de la distribution de 50 malades selon le sexe. Parmi ces 50 malades, 15 étaient de sexe masculin et 35 de sexe féminin. Le tableau d'effectifs est le tableau 1.

Tableau 1 : **Distribution de 50 malades selon le sexe**

Sexe	Effectif	Fréquence Relative	%
Masculin	15	0,30	30,0
Féminin	35	0,70	70,0
Total	50	1	100,0

Il faut noter que la fréquence relative est généralement exprimée en pourcentage.

2.2- MISE EN ORDRE DES DONNEES D'UN CARACTERE QUANTITATIF

2.2.1- Caractère quantitatif discontinu :

Si les modalités sont peu nombreuses, par exemple le nombre d'enfants dans une famille, on peut les laisser telles quelles, et on opérera de la même façon que dans le cas d'une variable qualitative.

On a observé, parmi 19 personnes, le nombre d'épisodes de syndrome grippal pendant un an. Le nombre d'épisodes variant de 0 à 4, on classe les 19 individus selon les valeurs entières successives 0, 1, 2, 3, 4.

Le tableau d'effectif correspondant est le tableau 2.

Tableau 2 : Distribution du nombre d'épisode de syndrome grippal parmi 19 personnes

Nombre D'épisodes	Effectif	%
0	3	15,8
1	7	36,8
2	6	31,6
3	2	10,5
4	1	5,3
Total	19	100,0

On peut aussi réunir les deux dernières modalités sous une rubrique unique «3 et plus ». L'effectif correspondant à cette modalité est $2 + 1 = 3$.

Si les modalités du caractère quantitatif discontinu sont très nombreuses, comme le nombre de globules rouges, on opérera de la même façon que dans le cas d'un caractère quantitatif continu.

2.2.2.- Caractère quantitatif continu :

Pour pouvoir dresser un tableau d'effectifs qui soit facile à lire, il faut grouper les observations ou données dans un certain nombre de classes successives, contiguës ne se recouvrant pas.

Par exemple, on a relevé le poids de 19 étudiants. L'unité de poids retenue est le Kilogramme et les résultats sont les suivants :

- 76,340 - 60,400 - 68,280 - 57,740 - 64,990
 - 83,450 - 79,650 - 64,100 - 72,880 - 69,120
 - 59,790 - 61,820 - 61,820 - 76,360 - 66,330
 - 52,990 - 70,560 - 70,130 - 65,450

Il s'agit de diviser le domaine de variation de la variable, de 52, 990 kg à 83, 450 kg, en classes.

Le tableau d'effectifs peut être présenté comme suit lorsqu'on retient 7 classes (tableau 3).

Tableau 3 : Répartition du poids de 19 étudiants

Classe N°	Poids (kg)	Effectif
1	50,000 à moins de 55,000	1
2	55,000 à moins de 60,000	2
3	60,000 à moins de 65,000	5
4	65,000 à moins de 70,000	4
5	70,000 à moins de 75,000	3
6	75,000 à moins de 80,000	3
7	80,000 à moins de 85,000	1
Total		19

Chaque classe est définie par **ses limites, son amplitude** et sa **valeur centrale**.

a- Les limites de classes : doivent être bien précisées, les classes étant mutuellement exclusives. Ainsi, les limites de la première classe sont 50,000 kg et 55,000 kg. Les limites de la deuxième classe sont 55,000 kg et 60,000 kg, etc.... Chaque classe a donc sa limite inférieure et sa limite supérieure. Cette dernière n'est autre que la limite inférieure de la classe subséquente.

Une observation ne doit être située que dans une seule classe. C'est pour cela que la convention adoptée est de toujours inclure dans la classe la limite inférieure et donc de toujours exclure la limite supérieure. Par exemple, un individu pesant 64,999 kg ne peut appartenir qu'à la troisième classe tandis qu'un individu pesant exactement 65,000 kg sera inclus dans la quatrième classe.

Pour alléger la présentation, les classes ne sont ordinairement pas écrites comme dans le tableau précédent. La première classe sera écrite 50-54, mais il est important d'insister sur le fait qu'elle englobe exactement les mêmes valeurs que la classe 50,000 kg à moins de 55,000 kg. Le tableau d'effectifs correspondant est le tableau 4.

Tableau 4 : Répartition du poids de 19 étudiants

Poids (kg)	Centre de Classe	Effectif	%
50-54	52,5	1	5,3
55-59	57,5	2	10,5
60-64	62,5	5	26,3
65-69	67,5	4	21,1
70-74	72,5	3	15,8
75-79	77,5	3	15,8
80-84	82,5	1	5,3
Total		19	100,0

Quelque fois, aux extrémités d'un tableau d'effectifs, on observe des classes ouvertes. Cela aurait consisté à écrire dans notre exemple pour la première classe : inférieure à 55 kg (< 55 kg) et pour la dernière classe : supérieure ou égal à 80 kg (> = 80 kg). Cette présentation devrait être évitée, sauf nécessité.

b/L'amplitude de la classe : est la différence entre les limites inférieure et supérieure de la classe. Dans notre exemple, toutes les classes ont une amplitude égale à 5 kg.

Souvent, on s'efforce de construire des classes d'amplitude égale. Il n'existe pas de règle permettant d'imposer le nombre de classe. Mais si on choisit une amplitude égale pour les classes, la règle suivante a pu être proposée :

$$c/\text{Nombre de classes} = \text{étendue} / \text{amplitude}$$

ou bien on peut commencer par calculer d'abord le nombre de classes selon cette formule,

$$nc = 1 + 10/3 * \log_{10}(n)$$

et par la suite on détermine l'amplitude : Amplitude = **étendu / nombre de classes**

d/L'étendue (ou marge) : est la différence entre la valeur la plus grande et la valeur la plus faible de la série.

L'étendue dans l'exemple du poids est :

$$83,450 - 52,990 = 30,460 \text{ kg}$$

Si on décide de prendre une amplitude de 5 kg, le nombre de classes est environ de : $30,46/5 = 6$ classes.

C'est ainsi que nous avons pris 7 classes. Il est intéressant de remarquer qu'il n'est pas obligatoire que la limite inférieure de la première classe soit 52 kg.

e/Le centre de classe : est la valeur située au milieu de la classe, c'est à dire à égale distance des limites inférieures et supérieure de la classe.

En d'autres termes, le centre de classe est égal :

Limite inférieure de la classe + limite supérieure de la classe /2

Le centre de la première classe, dans l'exemple du poids, est évidemment 52,5kg, celui de la deuxième classe 57,5 kg, etc....

Dans une classe, on ne distingue pas les valeurs qui y sont comprises. Toutes les valeurs de la classe sont confondues avec celle du centre de classe. On comprend alors que le fait de regrouper des valeurs en classe conduit à une perte d'information et donc de précision. Mais quand les différentes classes sont assimilées à des modalités d'une variable qualitative, l'analyse des données n'est pas compromise outre mesure.

Enfin pour les classes ouvertes, on ne peut pas calculer le centre de classe.

f/fréquences cumulées ou cumulatives :

Les fréquences cumulées peuvent donner de manière instantanée des informations très utiles. Elles se calculent aussi bien pour les effectifs que pour les fréquences relatives.

Pour les fréquences cumulées « **moins de** » ou « **croissantes** » on fait un cumul descendant à partir de la première classe : les fréquences sont additionnés au fur et à mesure. Les fréquences cumulées « moins de » permettent de dire combien il y a d'individus dont la valeur de la variable étudiée est inférieure à une limite donnée.

Reprenons l'exemple du poids des 19 étudiants et calculons les Fréquences cumulées. Les résultats sont consignés au tableau 5.

Tableau 5 : Répartition du poids de 19 étudiants

Poids (kg)	Effectif	%	Effectifs	Fréquences	Effectifs	Fréquences
			Cumulés « Moins de »	Relatives Cumulées « Moins de »	cumulés « Plus de »	Relatives Cumulées « Plus de »
			Croissant	Croissant	Décroissant	Décroissant
50-54	1	5,3	1	5,3	19	100
55-59	2	10,5	3	15,8	18	94,7
60-64	5	26,3	8	42,1	16	84,2
65-69	4	21,0	12	63,1	11	57,9
70-74	3	15,8	15	78,9	7	36,8
75-79	3	15,8	18	94,7	4	21,1
80-84	1	5,3	19	100,0	1	5,3
Total	19	100,0		-	-	-

On voit immédiatement que 12 étudiants ont un poids inférieur à 70 kg, soit 63,1 % des étudiants.

On aura noté que les fréquences relatives cumulées sont plus expressives que les effectifs cumulés. Pour cette raison, ce sont les fréquences relatives qui seront utilisées de préférence.

Les fréquences cumulées sont aussi utilisées pour la détermination des quantiles) et la construction des graphiques.

Soit X le caractère étudié,

-n la taille de l'échantillon,

-k le nombre de modalités du caractère X,

- x_1, \dots, x_k les modalités d'effectifs respectifs n_1, \dots, n_k .

-Fréquence de la modalité x_i est $f_i = n_i/n$

-Pourcentage de la modalité x_i est $p_i = n_i/n \cdot 100$

-Effectif cumulé croissant à la modalité x_i est $n_{croissant} = n_1 + n_2 + \dots + n_i$.

-Effectif cumulé décroissant à la modalité x_i est $n_{decroissant} = n_k + n_{k-1} + \dots + n_i$.

3-PRESENTATION TABULAIRE DE L'INFORMATION :

3.1.- DEFINITION :

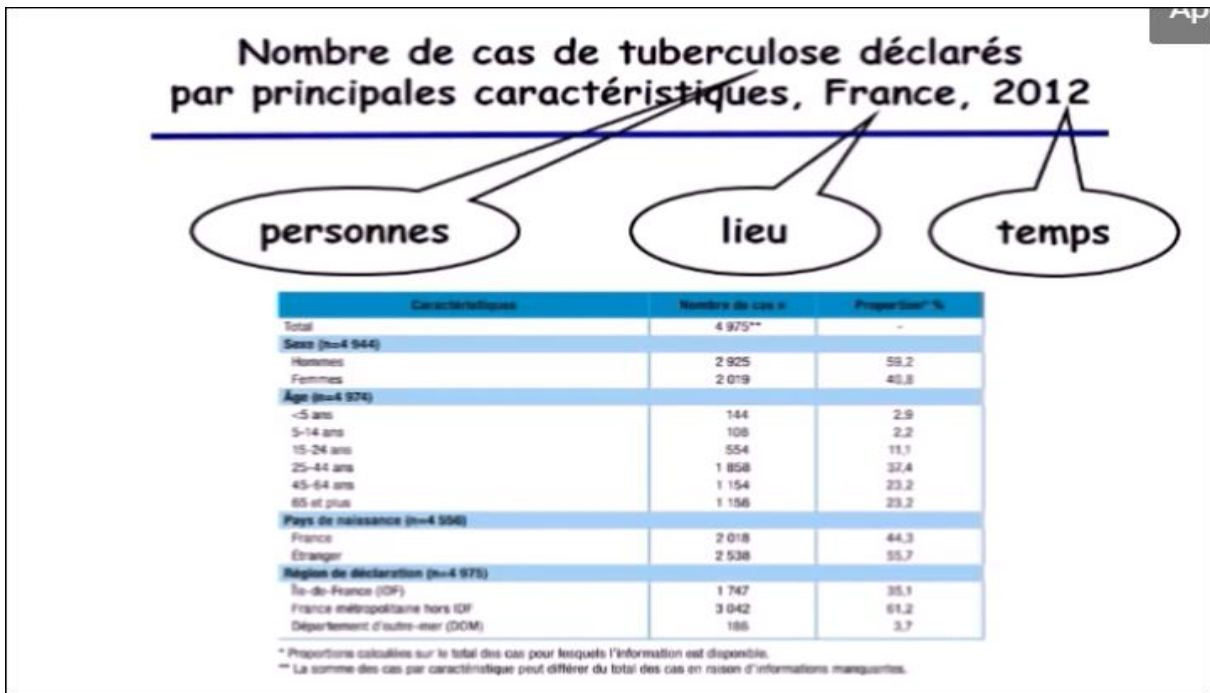
Un tableau est une unité d'information se suffisant à elle-même.

C'est pourquoi il doit contenir toutes les indications utiles à sa compréhension. Le lecteur doit être en mesure de cerner l'essentiel du message véhiculé par le tableau sans le recours du texte dans lequel néanmoins il doit être commenté.

3.2.- CONTENU DU TABLEAU :

Chaque tableau a son titre, son numéro et les notes qui l'accompagnent.

Le titre précise l'objet et définit le contenu du tableau. Le titre doit être explicite quand à l'objet et les éléments de l'étude, au lieu où se déroule la collecte des données et à la période de l'enquête. En d'autres termes, il doit répondre à la question : « quoi, où, quand ? ».



Le numéro est en chiffres arabes ou romains. Le numéro et le titre sont sur la même ligne au-dessus du tableau.

Des indicateurs complémentaires qui peuvent être indépendantes du titre ; sont portées en notes au bas du tableau. Il peut s'agir de la source des données, des unités de mesure, d'explications complémentaires ou de statistiques qui testent la liaison entre les variables, etc.

Le tableau proprement dit comporte :

- des têtes de colonnes avec un titre,
- des têtes de lignes avec un titre,
- un corps.

L'ossature d'un tableau se présente donc comme suit :

	Têtes de colonnes
Têtes de lignes	Corps Matrice Inscription de données chiffrées

Dans la présentation typographique moderne, les traits de séparation horizontaux et verticaux ont tendance à disparaître. Les cases sont virtuelles, la lecture en devient plus agréable et plus facile.

Les 5 principales localisations du cancer chez la femme en Algérie –Année 2016			
Localisation	Nombre de nouveaux cas	Taux brut /100 000	Age médian
Sein	9862	49,3	47
Colon rectum	2388	12,9	65
Thyroïde	1407	9,3	42
Col utérin	1092	7,2	62
Estomac	983	5,9	65
Total	14942 (64%)		

Source : Réseau National des Registres du cancer

Conclusion :

Tableau exemplaire :

- clair**, qui peut être lu et compris par le plus grand nombre avec
- des **données honnêtes**.

Références bibliographiques :

-Abdeljalil BEZZAOUCHA. Epidémiologie et bio statistiques à l'usage des étudiants en sciences médicales .Office des Publications Universitaires OPU 1996.

-Thierry ANCELLE . Statistique/Epidémiologie. Editions Maloine 2002.

-Thierry ANCELLE. 2014 (YouTube <https://www.youtube.com/c/ThierryAncellecelle> consulté le Novembre 2016)