

Les statistiques descriptives
Troisième Partie
Les tableaux statistiques

Emmanuel LESUEUR

Sommaire

1	Présentation et mise en forme des tableaux statistiques.....	2
2	Trier, Filtrer, Dédoublonner, Dénombrer sous Excel	4
3	Les tableaux à un caractère	5
3.1	Cas des caractères qualitatifs.....	6
3.2	Cas des caractères quantitatifs	7
4	Les tableaux à 2 caractères.....	8

1 Présentation et mise en forme des tableaux statistiques

Nous avons vu qu'un des objectifs de la statistique descriptive est de résumer les données «brutes» recueillies sur une population dans des tableaux statistiques. Cela permet de rendre les données lisibles. On présentera en ligne les modalités des caractères étudiés pour chaque individu et en colonne les caractères étudiés.

Par exemple, je réalise une enquête d'opinion sur l'accueil au PULDV¹ auprès de 11 étudiants de 1^{ère} année à l'EMLV².

Les critères (ou variables statistiques ou encore caractères) étudiés sont le Nom, Prénom, Age, Série du Bac, Opinion sur l'accueil au PULDV.

La matrice des données recueillies :

```
Prénom;Nom;Age;Série du bac;Opinion sur l'accueil au PULDV
Jacques;OUCHEDEMAIN;20;L;Très bon
Jean;BANBOIS;18;S;Très Très bon
Aude;JAVEL;19;L;Mauvais
Thomas;OUOKE;19;ES;Bon
Maguy;TARESTNAZ;20;STMG;Médiocre
Cécile;OUINKSA;17;S;Très bon
Anna;CONDA;17;ES;Bon
Mehdi;TÉRANNÉ;18;STMG;Bon
Sarah;MOLLIS;18;STMG;Très bon
Marshall;ESSENCE;17;ES;Médiocre
Marie;NOFARINGITE;18;ES;Bon
```

En tant que tel, la matrice de données ne peut pas être interprétée facilement. On présentera donc les résultats dans un tableau. Pour ce faire, nous utiliserons Excel pour créer et formater le tableau à partir de données brutes.

Résultat de l'enquête effectuée auprès d'étudiants de l'EMLV.

Prénom	Nom	Age	Série du Bac	Opinion sur l'accueil
Jacques	OUCHEDEMAIN	20	L	Très bon
Jean	BANBOIS	18	S	Très bon
Aude	JAVEL	19	L	Mauvais
Thomas	OUOKE	19	ES	Bon
Maguy	TARESTNAZ	20	STMG	Médiocre
Cécile	OUINKSA	17	S	Très bon
Anna	CONDA	17	ES	Bon
Mehdi	TÉRANNÉ	18	STMG	Bon
Sarah	MOLLIS	18	STMG	Très bon
Marshall	ESSENCE	17	ES	Médiocre
Marie	NOFARINGITE	18	ES	Bon

Figure 1 : Présentation lisible des résultats de l'enquête

Le tableau statistique quant à lui, doit mettre en évidence :

- ☞ La variable statistique,
- ☞ Ses modalités
- ☞ L'effectif correspondant à chaque modalité.

En reprenant l'exemple de l'étude ci-dessus,

- La **variable statistique** est l'opinion sur l'accueil :
- Les **modalités** sont Mauvais, Médiocre, Bon et Très bon.
- Les **effectifs** correspondants sont respectivement : 1, 2, 4 ; 4

¹ Pôle Universitaire Léonard De Vinci

² Ecole de Management Léonard de Vinci

Le tableau statistique correspondant se présentera ainsi :

Opinion sur l'accueil au PULDV	
Opinion sur l'accueil (x_i)	Effectif (n_i)
Mauvais	1
Médiocre	2
Bon	4
Très bon	4
Total	11

Figure 2 : Tableau statistique

Ainsi, d'une façon générale, pour construire un tableau statistique, il conviendra :

- De déterminer le caractère à étudier (x)
- D'identifier ses modalités (x_i)
- De calculer l'effectif correspondant pour chaque modalité. (n_i)

Le tableau statistique se présentera sous la forme suivante :

Modalité x_i	Effectif n_i
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_j	n_j
.	.
.	.
.	.
x_k	n_k

Figure 3 : Etude du caractère x

La 1^{ère} colonne du tableau reprend les différentes **modalités (x_i)** prises par la variable ou le caractère.

La 2^{ème} colonne présente les **effectifs (n_i)** (ou **fréquences absolues**), c'est-à-dire le nombre d'individus correspondant à chaque modalité du caractère.

Définition : L'effectif partiel (ou fréquence absolue) n_i d'une modalité x_i est le nombre d'individus ayant cette modalité.

Définition : L'effectif total (ou taille) d'une population, noté n , est le nombre d'individus qui composent cette population ; c'est donc la somme des effectifs partiels. On a donc:

$$\sum_{i=1}^k n_i = n$$

Définition : On appelle **distribution statistique** (ou série statistique), l'ensemble des modalités et des effectifs d'un caractère une que l'on note : $\{x_i, n_i\}$.

La présentation d'un tableau statistique doit respecter des principes généraux incontournables :

- ☞ Il doit porter un titre précisant son contenu (phénomène étudié, lieu, date, etc..).
- ☞ Il doit porter des intitulés de colonnes clairement définis.

- ☞ Il doit contenir les unités utilisées (€, k€, \$US, m², kg, etc...).
- ☞ Il faut indiquer la source des informations lorsque les données sont empruntées à une publication ou à un organisme.

Les tableaux statistiques peuvent avoir une ou à plusieurs dimensions.

À «une dimension» si un seul caractère est étudié (nombre d'enfants/ménage)

À «deux dimensions» si l'on étudie deux caractères (nombre et sexe des enfants/ménage)

À n dimensions si l'on retient n caractères (Attention à ne pas prendre un n trop grand pour pouvoir analyser les données !)

Application : Nous allons maintenant comprendre les formats de cellules sous Excel et apprendre à mettre en forme un tableau. Nous utiliserons le classeur « 3a-SD_Tableaux et Formats.xlsx »

2 Trier, Filtrer, Dédoublonner, Dénombrer sous Excel

Prenons maintenant l'exemple d'une enquête réalisée auprès d'un échantillon de 50 familles françaises sur le nombre d'enfants par ménage.

Données « brutes » : 3 4 0 2 0 1 1 5 2 5 4 1 8 4 1 1 2 7 2 5 2 4 2 2 6 2 2 3 9 3 7 3 0 3 3 4 3 4 2 4 5 2 2 6 6 6 7 7 8 9

Là encore, les résultats ne sont pas lisibles. Il faut les regrouper dans un tableau pour faciliter le traitement et les interprétations. Pour calculer les effectifs, il convient, dans un 1^{er} temps de regrouper par modalités les données brutes.

Données « brutes » : 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 6 6 6 7 2 8 9

On constate que le critère *Nombre d'enfants* comporte 10 modalités, allant de 0 à 9. Pour chacune d'entre elle, on calculera l'effectif, pour obtenir la série brute suivante :

Nombre d'enfants;Effectif

0 ;3 1 ;5 2 ;14 3 ;9 4 ;7 5 ;5 6 ;3 7 ;2 8 ;1 9 ;1

Et obtenir le tableau de statistique suivant :

Nombre d'enfants par famille observé dans un échantillon de 50 familles

Nombre d'enfants par famille (x_i)	Effectif (n_i)
0	3
1	5
2	14
3	9
4	7
5	5
6	3
7	2
8	1
9	1
Total	50

Modalités x_i prises par la variable x

Nombre de ménages correspondant à chaque modalité

Figure 4 : Tableau de statistiques sous Excel (Ex 2)

Application : Nous allons maintenant apprendre à réaliser ses différentes opérations sous Excel et à calculer automatiquement les effectifs.

Nous utiliserons pour cela, le classeur « 3b-SD_Trier_Filter_Dédoublonner_Dénombrer.xlsx ».

3 Les tableaux à un caractère

On considère une population statistique de n individus décrite selon le caractère x dont les k modalités sont $x_1, x_2, \dots, x_i, \dots, x_k$

Définition : On appelle fréquence (ou fréquence relative) f_i , la proportion des individus présentant la même modalité dans la population étudiée. Elle est obtenue calculant le ratio de chaque effectif et l'effectif total. On a donc :

$$f_i = \frac{n_i}{n}$$

La somme des fréquences relatives est égale = 1.

C'est une grandeur sans unité qui peut donc s'exprimer en pourcentage :

$$f_i(\%) = \frac{n_i}{n} * 100^1$$

La somme des fréquences relatives exprimées en pourcentage est égale = 100.

D'une façon générale, le tableau statistique s'écrira ainsi :

Modalité x_i	Effectif n_i	Fréquence f_i	f_i (%)
x_1	n_1	f_1	f'_1
x_2	n_2	f_2	f'_2
.	.	.	.
.	.	.	.
.	.	.	.
x_j	n_j		f'_j
.	.	.	.
.	.	.	.
.	.	.	.
x_k	n_k	f_k	f'_k
Total	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k f_i = 1$	$\sum_{i=1}^k f'_i = 100$

Figure 5 : Etude du caractère x (Ex 2)

Reprenons l'exemple du nombre d'enfants par famille pour calculer les fréquences relatives en valeur et en pourcentage.

Nombre d'enfants par ménage relevé dans un échantillon de 50 familles françaises

Nbre d'enfants par famille (x_i)	Effectif (Fréquence absolue) (n_i)	Effectif partiel (f_i) (Fréquence relative)	
		Valeur	Pourcentage
0	3	0,06	6,00%
1	5	0,1	10,00%
2	14	0,28	28,00%
3	9	0,18	18,00%
4	7	0,14	14,00%
5	5	0,1	10,00%
6	3	0,06	6,00%
7	2	0,04	4,00%
8	1	0,02	2,00%
9	1	0,02	2,00%
Total	50	1	100,00%

Figure 6 : Calculs des effectifs partiels en valeur et en % (Ex 1)

Prenons un autre exemple : Le service qualité d'une entreprise compte le nombre de pièces défectueuses produites par les 20 machines qui produisent la même pièce durant un mois.

¹ Sous Excel, on ne fera pas la multiplication par 100, on appliquera le format % !

Nombre de pièces défectueuses produites
sur 20 machines pendant 1 mois

Nombre de pièces défectueuses (x_i)	Effectif (Fréquence absolue) (n_i)	Effectif partiel (f_i) (Fréquence relative)	
		Valeur	Pourcentage
3	4	0,2	20,00%
5	2	0,1	10,00%
7	3	0,15	15,00%
8	3	0,15	15,00%
9	3	0,15	15,00%
10	1	0,05	5,00%
14	2	0,1	10,00%
16	1	0,05	5,00%
33	1	0,05	5,00%
Total	20	1	100,00%

Figure 7 : Calculs des effectifs partiels en valeur et en % (Ex 2)

Voyons comment réaliser de tels calculs avec Excel. Pour cela, il convient de comprendre les références relatives et absolues sous Excel. Nous utilisons le classeur :

«3c-SD_References_absolues.xlsx »

3.1 Cas des caractères qualitatifs

Nous avons vu que les critères qualitatifs peuvent avoir des modalités nominales ou ordinales. Voyons le cas des tableaux statistiques dont les caractères sont à **modalités nominales**. Prenons le cas de la répartition des salariés de la société « HaNoNyme » selon leur CSP¹ au 31/12/2017. Le tableau statistique correspondant sera sous la forme :

CSP (x_i)	Effectif (n_i)	Fréquence (f_i)	
		Valeur	%
Cadres Supérieurs	15	0,07177	7,18%
Agent de maîtrise	6	0,02871	2,87%
Employés	40	0,19139	19,14%
Ouvriers spécialisés	48	0,22967	22,97%
Ouvriers qualifiés	92	0,44019	44,02%
Autres	8	0,03828	3,83%
Total	209	1	100,00%

Figure 8 : Répartition des salariés de la société « HaNoNyme »
selon leur CSP au 31/12/2017

Voyons maintenant le cas des tableaux statistiques dont les caractères sont à **modalités ordinales**. Reprenons l'exemple de l'enquête d'opinion sur l'accueil au PULDV auprès de 11 étudiants de 1^{ère} année à l'EMLV.

Prénom	Nom	Age	Série du Bac	Opinion sur l'accueil
Jacques	OUCHEDMAIN	20	L	Très bon
Jean	BANBOIS	18	S	Très bon
Aude	JAVEL	19	L	Mauvais
Thomas	OUOKE	19	ES	Bon
Maguy	TARESTNAZ	20	STMG	Médiocre
Cécile	OUINKSA	17	S	Très bon
Anna	CONDA	17	ES	Bon
Mehdi	TÉRANNÉ	18	STMG	Bon
Sarah	MOLLIS	18	STMG	Très bon
Marshall	ESSENCE	17	ES	Médiocre
Marie	NOFARINGITE	18	ES	Bon

Figure 9 : Résultat de l'enquête d'opinion sur l'accueil au PULDV

¹ Catégorie Socio-Professionnelle

Il convient de déterminer et d'ordonner les modalités du caractère « Opinion sur l'accueil », soit : « Très bon », « Bon », « Médiocre », « Mauvais ». Ensuite, il faut calculer les effectifs pour chaque modalité. Enfin, on calculera les fréquences en valeur et en pourcentage. Le tableau suivant résume la série statistique :

Opinion sur l'accueil au PULDV

Opinion sur l'accueil (x_i)	Effectif (n_i)	Fréquence (f_i)	
		Valeur	%
Très bon	4	0,36364	36,36%
Bon	4	0,36364	36,36%
Médiocre	2	0,18182	18,18%
Mauvais	1	0,09091	9,09%
Total	11	1	100,00%

Figure 10 : Répartition des étudiants selon leur opinion sur l'accueil au PULDV

3.2 Cas des caractères quantitatifs

Les tableaux à caractères quantitatifs peuvent contenir plus d'informations que les tableaux à caractères qualitatifs ; on pourra notamment calculer :

- ☞ Les effectifs cumulés
- ☞ Les fréquences cumulées

Cela permettra de répondre à des questions du genre, « Combien de famille ont plus de 3 enfants ? », ou « Combien de famille ont moins de 5 enfants ? ». Et en terme de fréquence, « Quelle est la proportion des familles ayant plus de 3 enfants ? » ou « Quelle est la part des familles ayant moins de 5 enfants ? ».

Pour calculer un effectif cumulé (noté $N(x)$), il suffit d'ajouter à l'effectif d'une modalité d'un caractère, l'effectif de la modalité précédente. Il en est de même pour déterminer les fréquences cumulées (notée $F(x)$). On distinguera les cumuls « ascendants » des cumuls « descendants » selon que la somme se fera respectivement de haut en bas ou de bas en haut.

A. Cas des critères quantitatifs discrets.

En reprenant l'exemple du nombre d'enfants par ménage dans un échantillon de 50 familles, et à l'aide du tableur Excel on obtient le tableau ci-dessous :

Nbre d'enfants par famille (x_i)	Effectif (n_i)	Fréquence (f_i)	Effectifs cumulés		Fréquences cumulées	
			"Moins de"	"Plus de"	"Moins de"	"Plus de"
0	3	0,06	3	50	6%	100%
1	5	0,1	8	47	16%	94%
2	14	0,28	22	42	44%	84%
3	9	0,18	31	28	62%	56%
4	7	0,14	38	19	76%	38%
5	5	0,1	43	12	86%	24%
6	3	0,06	46	7	92%	14%
7	2	0,04	48	4	96%	8%
8	1	0,02	49	2	98%	4%
9	1	0,02	50	1	100%	2%
Total	50	1	****	****	****	****

Figure 11 : Nombre d'enfants (x_i) observés dans un échantillon de 50 familles

Interprétation :

De cette étude, nous pouvons dire que :

22 ménages de l'échantillon ont « moins de » 3 enfants. On peut dire aussi qu'il y a 22 ménages dans la population qui ont 2 enfants maximum.

12 ménages de l'échantillon ont «plus de» 4 enfants. On peut dire aussi qu'il y a 12 ménages dans l'échantillon qui ont «au moins» 5 enfants.

76% des ménages de la population ont «moins de» 5 enfants ou 4 enfant maximum.

8% des ménages de la population ont «plus de» 6 enfants ou «au moins» 5 enfants.

B. Cas des critères quantitatifs continus.

Prenons l'exemple des salaires de la société « HaNoNyme » au 31/12/2017. Sous Excel, on obtient le tableau statistique suivant :

Classe de salaires mensuels bruts (€)	Effectif (n_i)	Fréquence (f_i)	Effectifs cumulés		Fréquences cumulées	
			"Moins de"	"Plus de"	"Moins de"	"Plus de"
[1500-2000[55	0,263	55	209	26%	100%
[2000-2500[29	0,139	84	154	40%	74%
[2500-3000[25	0,120	109	125	52%	60%
[3000-3500[26	0,124	135	100	65%	48%
[3500-4000[36	0,172	171	74	82%	35%
[4000-4500[23	0,110	194	38	93%	18%
[4500-5000[7	0,033	201	15	96%	7%
[5000-10000[5	0,024	206	8	99%	4%
10000 et +	3	0,014	209	3	100%	1%
Total	209	1	-----	-----	-----	-----

Figure 12 : Répartition des salaires de la société « HaNoNyme » au 31/12/2017

Interprétation :

125 salariés touchent plus de 2500€, soit 60%.

82% des personnes touchent moins de 4000€, soit 171 salariés.

201 personnes ont un salaire strictement inférieur à 5000€, soit 96%.

7% des salariés touchent plus que 4500€, soit 15 personnes.

C Application sous Excel

Réaliser sous Excel les 2 tableaux précédents (Figure 11, page 7 et Figure 12, page 8) en utilisant le classeur « 3d-Effectifs et fréquences cumulées.xlsx », feuilles « Critères quantitatifs discrets » et « Critères quantitatifs continus »).

4 Les tableaux à 2 caractères

Une population statistique peut être décrite à l'aide de 2 caractères simultanément.

Exemple 1 : La population des ménages peut être décrite selon son **revenu** et ses **dépenses**.

Exemple 2 : La population active française peut être décrite en fonction de la **CSP** et du **niveau de formation**.

Les tableaux statistiques correspondants sont donc à 2 dimensions. On les appelle tableaux de contingence ou à tableaux à double entrée. Comme pour les tableaux à un caractère, le tableau de contingence répond à une notation conventionnelle :

En colonne, on indiquera les n modalités $x_1, x_2, x_i, \dots, x_n$ du caractère X

En ligne, on trouvera les k modalités $y_1, y_2, y_j, \dots, y_k$ du caractère Y

L'effectif n_{ij} correspond à l'intersection d'une ligne i et d'une colonne j .

Pour mieux comprendre, prenons l'exemple de la répartition des salaires de la société « HaNoNyme » selon le genre (x_i) et le niveau de formation (y_i). Le tableau statistique sera le suivant :

$Y_j \backslash X_i$	Homme	Femme	Total ($n_{i.}$)
BEP-CAP	80	12	92
BAC	59	29	88
Licence	2	3	5
Maîtrise	13	9	22
Doctorat	1	1	2
Total ($n_{.j}$)	155	54	209

Distribution marginale
des salairés selon le genre
↓

Distribution marginale
des salairés selon le
niveau de formation →

Figure 13 : Répartition des salaires de la société « HaNoNyme » selon le genre et le niveau de formation au 31/12/2017