

Statistique descriptive

Chapitre 2 : Le tableau statistique

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Le premier résumé	2
De 500 lignes à quelques lignes	2
Le dépouillement, à la main	2
Les cinq colonnes	2
L'anatomie du tableau	2
Effectif et fréquence	3
Pourquoi les deux colonnes	3
Les cumuls	3
Le cumul répond aux questions de seuil	4
Le tableau, en pratique	4
Une variable nominale	4
Une variable ordinale	4
Une variable quantitative discrète	5
Le cumul en escalier	5
Trois questions, trois calculs	6
Les erreurs classiques	6
Arrondis et non-réponses	6
Le pourcentage sans base	6
S'entraîner	7
Exercice 1 : un tableau à trous	7
Exercice 1 : solution	7
Exercice 2 : deux agences, deux tailles	7
Exercice 2 : solution	8
Ce qu'il faut retenir	8
Le résumé du chapitre	8

Le premier résumé

De 500 lignes à quelques lignes

Ce qu'on cherche à faire

Les données brutes sont une liste : salarié 1, salarié 2, ... salarié 500. Elle contient toute l'information et n'en délivre aucune.

Le **tableau statistique** regroupe les individus qui partagent la même modalité et ne retient que les **comptages**. On perd l'identité des individus ; on gagne une vue d'ensemble.

Le principe

On passe d'une liste de n individus à un tableau de k lignes, où k est le nombre de modalités — et k est toujours **beaucoup plus petit** que n .

Le dépouillement, à la main

Le diplôme des 20 premiers salariés du fichier

Bac+5, Bac, Bac, Bac+2, Bac+5, Bac+2, Bac, Bac+2, Bac+5, Bac+5, Bac, Bac+5, Bac+2, Bac+5, Bac+5, Bac, Bac+5, Bac, Bac+2, Bac+5.

On compte chaque modalité, puis on vérifie que le total redonne 20 :

Diplôme	Comptage	n_i
Bac	IIIII I	6
Bac+2	IIIII	5
Bac+5	IIIII IIII	9
Total		20

Vingt lignes ne sont pas cinq cents

Ici 45 % de Bac+5, contre 21 % dans l'ensemble du fichier : les vingt premières lignes ne résument pas l'entreprise. Le **tableau décrit les individus comptés, et eux seuls**.

Les cinq colonnes

L'anatomie du tableau

le **tableau statistique** : tout ce qu'on peut dire d'une variable, en cinq colonnes

x_i	n_i	f_i	N_i	F_i
x_1	n_1	f_1	N_1	F_1
x_2	n_2	f_2	N_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k	$N_k = n$	$F_k = 1$
total	n	1		

x_i les *modalités* ou les classes

n_i l'*effectif* : combien d'individus

$f_i = n_i/n$ la *fréquence* : quelle part

N_i l'*effectif cumulé* : combien au plus

F_i la *fréquence cumulée*

Chaque colonne appelle son propre graphique

— c'est le chapitre 5.

Effectif et fréquence

Effectif

L'**effectif** n_i de la modalité x_i est le nombre d'individus qui la possèdent.

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

Fréquence

La **fréquence** f_i est la part relative de la modalité :

$$f_i = \frac{n_i}{n} \quad \text{avec} \quad 0 \leq f_i \leq 1 \quad \text{et} \quad \sum_{i=1}^k f_i = 1$$

Multipliée par 100, elle donne un **pourcentage**.

Pourquoi les deux colonnes

Combien, et quelle part

L'effectif dit **combien**, la fréquence dit **quelle part**.

L'effectif est indispensable pour juger de la fiabilité : une modalité à 80 % calculée sur 5 individus ne vaut rien. La fréquence est indispensable pour **comparer** deux populations de tailles différentes.

Les deux vérifications systématiques

La somme des effectifs vaut n , la somme des fréquences vaut 1. Ces deux contrôles détectent la quasi-totalité des erreurs de saisie.

Les cumuls

Effectif et fréquence cumulés croissants

$$N_i = \sum_{j \leq i} n_j \quad F_i = \sum_{j \leq i} f_j = \frac{N_i}{n}$$

N_i répond à : combien d'individus ont une valeur **inférieure ou égale** à x_i ?

Cumuls décroissants

$$N_i^\downarrow = n - N_{i-1} \quad F_i^\downarrow = 1 - F_{i-1}$$

Ils répondent à : combien d'individus ont une valeur **supérieure ou égale** à x_i ?

Le cumul répond aux questions de seuil

La colonne la plus utile en pratique

Les questions d'un décideur ne portent presque jamais sur une modalité isolée. Elles portent sur des **seuils** :

- combien de salariés gagnent moins de 6 000 dirhams ?
- quelle part de l'effectif dépasse dix ans d'ancienneté ?
- au-dessous de quel salaire se situe la moitié du personnel ?

Ces trois questions se lisent directement dans les colonnes cumulées. La troisième donne la **médiane**, et c'est ainsi qu'on la calculera au chapitre 6.

Rappel du chapitre 1

Cumuler suppose un **ordre**. La colonne N_i n'a donc aucun sens pour une variable nominale : cumuler « Tanger-Tétouan » et « Casablanca » ne veut rien dire. Elle a un sens pour une variable ordinale, discrète ou continue.

Le tableau, en pratique

Une variable nominale

La région de nos 500 salariés

Région	n_i	f_i
Tanger-Tétouan	180	0,360
Casablanca	137	0,274
Rabat	102	0,204
Autres	81	0,162
Total	500	1,000

Ni cumul, ni ordre imposé

Pas de colonne N_i : la région n'a pas d'ordre. On range alors les modalités par **effectif décroissant**, la modalité « Autres » en dernier : c'est la présentation qui se lit le plus vite.

Une variable ordinale

Le diplôme de nos 500 salariés

x_i	n_i	f_i	N_i	F_i
Bac	219	0,438	219	0,438
Bac+2	174	0,348	393	0,786
Bac+5	107	0,214	500	1,000
Total	500	1,000		

Ce qu'on lit en trois secondes

Un peu moins de la moitié des salariés s'arrête au baccalauréat ; un sur cinq environ est titulaire d'un Bac+5. Les cumuls sont licites parce que le diplôme est **ordinal** : 78,6 % des salariés ont **au plus** un Bac+2.

Une variable quantitative discrète

L'ancienneté, valeurs groupées

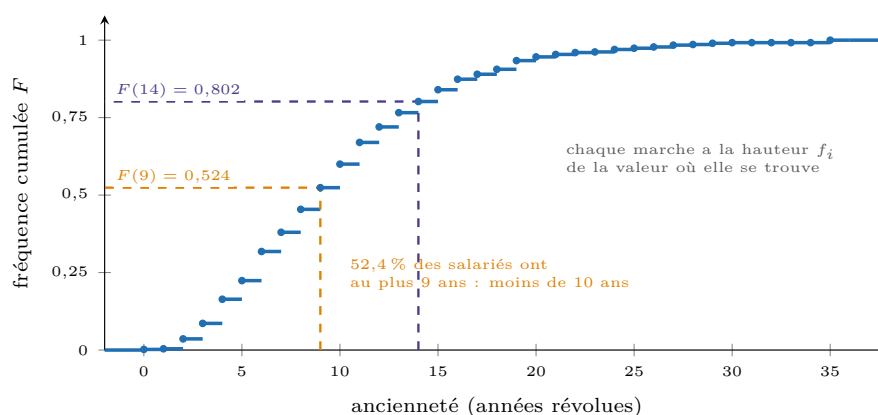
ancienneté	n_i	f_i	N_i	F_i
0 à 4 ans	82	0,164	82	0,164
5 à 9 ans	180	0,360	262	0,524
10 à 14 ans	139	0,278	401	0,802
15 à 19 ans	66	0,132	467	0,934
20 ans et plus	33	0,066	500	1,000
Total	500	1,000		

La lecture par seuils

Plus de la moitié du personnel — 52,4 % — a moins de dix ans d'ancienneté ; un salarié sur quinze seulement en a vingt ou davantage. L'entreprise a beaucoup recruté récemment, et sa mémoire interne est concentrée sur peu de personnes.

Le cumul en escalier

la fréquence cumulée de l'ancienneté : une fonction en escalier



Une fonction, pas seulement une colonne

F monte d'une **marche** à chaque valeur observée et reste plate entre deux : on y lit tous les seuils à la fois.

Trois questions, trois calculs

Avec les colonnes cumulées de l'ancienneté

Question	Calcul	Réponse
Moins de 10 ans ?	$F(9)$	52,4 %
Au moins 15 ans ?	$1 - F(14)$	19,8 %
De 5 à 14 ans ?	$F(14) - F(4)$	63,8 %

Les cumuls décroissants, sur le même tableau

Au moins	0 an	5 ans	10 ans	15 ans	20 ans
$N^\downarrow$	500	418	238	99	33
$F^\downarrow$	1	0,836	0,476	0,198	0,066

Chaque $F^\downarrow$ vaut 1 moins le F de la ligne précédente : $0,476 = 1 - 0,524$.

Les erreurs classiques

Arrondis et non-réponses

Piège 1 : les fréquences qui ne somment pas à un

Presque toujours un **arrondi**. $0,438 + 0,348 + 0,214 = 1,000$ ici, mais avec trois modalités à 33,3 % on obtient 99,9 %.

Ne jamais forcer le total. On signale l'arrondi en note de bas de tableau : c'est la pratique de tous les instituts statistiques.

Piège 2 : les non-réponses

Si 30 salariés n'ont pas indiqué leur diplôme, faut-il diviser par 500 ou par 470 ? Les deux se pratiquent, mais **il faut le dire** : fréquence **brute** sur $n = 500$, ou **nette**, hors non-réponse, sur $n = 470$. Un écart de 6 % sur le dénominateur est un écart de 6 % sur tous les résultats.

Le pourcentage sans base

Piège 3 : le pourcentage sans base

« Les ventes ont progressé de 200 % » : de 1 à 3 unités, ou de 1 million à 3 millions ?

Un pourcentage sans son effectif n'est pas une information. C'est le procédé rhétorique le plus courant dans la presse économique.

La règle professionnelle

Un tableau publié porte toujours, sous le titre :

- le **champ** (quelle population, quelle date) ;
- l'**effectif total** n ;

- la **source** ;
- la mention des non-réponses s'il y en a.

Sans ces quatre éléments, le tableau n'est pas exploitable — même si les chiffres sont exacts.

S'entraîner

Exercice 1 : un tableau à trous

Le nombre de pièces de 200 logements

Compléter le tableau, en justifiant chaque case.

Pièces x_i	n_i	f_i	N_i	F_i
1	16			
2		0,26		
3				0,66
4	44			
5 et plus				
Total	200			

Puis : quelle part des logements a **au moins trois** pièces ?

Exercice 1 : solution

Le tableau complété

Pièces x_i	n_i	f_i	N_i	F_i
1	16	0,08	16	0,08
2	52	0,26	68	0,34
3	64	0,32	132	0,66
4	44	0,22	176	0,88
5 et plus	24	0,12	200	1,00
Total	200	1,00		

La démarche

$n_2 = 0,26 \times 200 = 52$; $N_3 = 0,66 \times 200 = 132$, d'où $n_3 = 132 - 68 = 64$; le dernier effectif complète à 200 : $n_5 = 200 - 176 = 24$. Au moins trois pièces : $1 - F_2 = 1 - 0,34 = 66\%$.

Exercice 2 : deux agences, deux tailles

La satisfaction des clients

Une banque interroge les clients de deux agences.

Satisfaction	Agence de Tétouan	Agence de Casablanca
insatisfait	6	72
neutre	10	108
satisfait	24	180
Total	40	360

1. Quelle agence compte le plus de clients satisfaits ? Est-ce la bonne question ?
2. Construire, pour chaque agence, les colonnes f_i et F_i .
3. Dans quelle agence la part des clients **au plus neutres** est-elle la plus forte ?
4. Quelle réserve faut-il formuler sur l'agence de Tétouan ?

Exercice 2 : solution

Les fréquences et leurs cumuls

	f_i Tétouan	F_i Tétouan	f_i Casa.	F_i Casa.
insatisfait	0,150	0,150	0,200	0,200
neutre	0,250	0,400	0,300	0,500
satisfait	0,600	1,000	0,500	1,000

Les réponses

1. Casablanca (180 contre 24), mais elle a neuf fois plus de clients : **on compare des fréquences**.
2. Voir le tableau : 60 % de satisfaits à Tétouan, 50 % à Casablanca.
3. Au plus neutres : $F_2 = 40\%$ à Tétouan, 50 % à Casablanca.
4. Tétouan repose sur 40 réponses : un client de plus déplace les pourcentages de plus de deux points. **L'effectif doit accompagner la fréquence**.

Ce qu'il faut retenir

Le résumé du chapitre

Six points

1. Le tableau statistique fait passer de n lignes à k lignes : c'est le premier résumé.
2. **Effectif** n_i : combien. **Fréquence** $f_i = n_i/n$: quelle part. Les deux sont nécessaires.
3. **Cumuls** N_i et F_i : ils répondent aux questions de **seuil**, et exigent un ordre.
4. Une part comprise entre deux seuils se lit par **différence** de deux F ; une part au-dessus d'un seuil, par $1 - F$.
5. Deux contrôles systématiques : $\sum n_i = n$ et $\sum f_i = 1$.
6. Un tableau se publie avec son **champ**, son **effectif** et sa **source**.

La suite

Pour une variable continue, on ne peut pas énumérer les modalités : il faut les **regrouper en classes**. Et ce regroupement obéit à des règles précises, dont dépend la validité de l'histogramme. C'est le chapitre 3.