

Statistique descriptive

Chapitre 4 : Une colonne, un graphique

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Le principe	2
Chaque colonne appelle son graphique	2
Les quatre graphiques	2
Vue d'ensemble	2
Le diagramme en bâtons	2
Dans quel ordre ranger les bâtons	3
Le diagramme à secteurs	3
Les limites du secteur	3
L'histogramme	4
Le polygone des fréquences	4
Trois groupes, trois polygones	4
Les courbes cumulatives	5
Comment les lire	5
Lire entre deux bornes	5
Lire la médiane sur la courbe	6
Bien faire un graphique	6
Les règles	6
L'axe tronqué, en image	7
Ni ombre, ni relief	7
S'entraîner	7
Exercice 1 : le budget d'un ménage	7
Exercice 1 : solution	8
Exercice 2 : lire une courbe dans les deux sens	8
Exercice 2 : solution	8
Ce qu'il faut retenir	9
Le résumé du chapitre	9

Le principe

Chaque colonne appelle son graphique

La correspondance

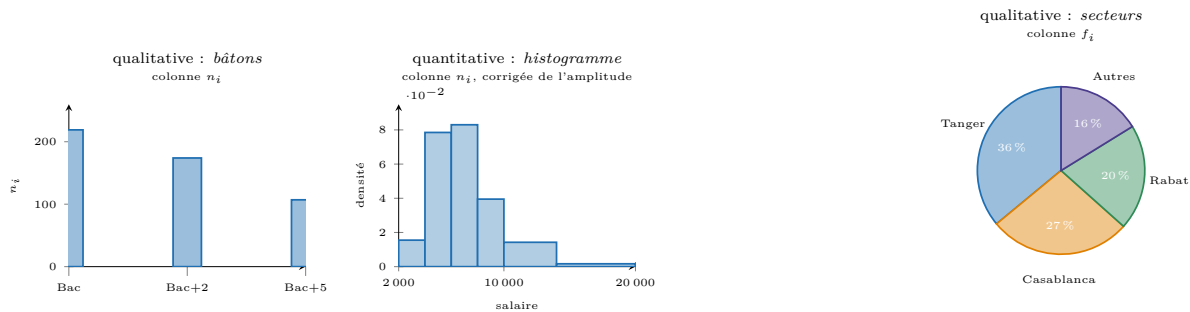
Colonne du tableau	Graphique	Type de variable
n_i	diagramme en bâtons	qualitative, discrète
f_i	diagramme à secteurs	qualitative
n_i corrigé de a_i	histogramme	continue
F_i	courbe cumulative	ordinaire, discrète, continue

Un graphique n'illustre pas : il montre

Un tableau donne les valeurs exactes ; un graphique donne la **forme**. Le tableau répond à « combien ? », le graphique à « à quoi cela ressemble-t-il ? » — l'asymétrie, deux bosses, une valeur aberrante. Ces choses-là ne se voient pas dans une colonne de chiffres.

Les quatre graphiques

Vue d'ensemble



Le diagramme en bâtons

Construction

Un **segment vertical** par modalité, de hauteur proportionnelle à l'effectif ou à la fréquence. Les bâtons sont **séparés**.

Pourquoi les bâtons sont séparés

L'espace entre deux bâtons signifie qu'il n'y a **rien** entre les deux modalités : aucune valeur intermédiaire n'existe. C'est l'inverse de l'histogramme, dont les rectangles sont **jointifs** parce que la variable prend toutes les valeurs de l'intervalle.

Ce détail visuel encode le type de la variable. Des barres collées pour une variable qualitative sont une faute de lecture, même si le graphique est joli.

Dans quel ordre ranger les bâtons

Le tri des modalités

- Variable **ordinaire** : on respecte l'ordre naturel — Bac, Bac+2, Bac+5. Jamais autre chose.
- Variable **nominale** : aucun ordre n'existe. On trie donc par effectif **décroissant** pour faciliter la lecture.

Trier une variable ordinaire par effectif détruit l'information la plus utile qu'elle contient.

Sur le fil rouge

Diplôme, ordre naturel	n_i	Région, par effectif	n_i
Bac	219	Tanger-Tétouan	180
Bac+2	174	Casablanca	137
Bac+5	107	Rabat	102
		Autres	81

Le diagramme à secteurs

Construction

Un disque découpé en secteurs, l'angle de chacun étant proportionnel à la fréquence :

$$\alpha_i = 360^\circ \times f_i$$

Sur la région

Région	n_i	f_i	angle
Tanger-Tétouan	180	0,360	129,6°
Casablanca	137	0,274	98,6°
Rabat	102	0,204	73,4°
Autres	81	0,162	58,3°
Total	500	1,000	360°

Les limites du secteur

Des limites réelles

Le secteur ne convient qu'aux **parts d'un tout**, donc à la colonne f_i , et seulement si :

- les modalités sont peu nombreuses — au-delà de cinq ou six, il devient illisible ;
- les parts sont nettement différentes — l'œil compare mal deux angles voisins.

Longueurs, angles, aires

L'œil humain compare bien des longueurs, mal des angles et très mal des aires. Rabat (73°) et « Autres » (58°) se départagent difficilement sur un disque ; sur des barres, l'écart de 21 salariées saute aux yeux. Le secteur ne garde qu'un avantage : il montre qu'on a

bien affaire à un tout.

L'histogramme

Construction

Des rectangles **jointifs**, dont la base est la classe et **l'aire** l'effectif. La hauteur est donc la densité $d_i = n_i/a_i$ (chapitre 3).

Ce qu'on lit sur l'histogramme des salaires

1. Le sommet — la **classe modale** — se situe entre 6 000 et 8 000 dirhams.
 2. La distribution est **étalée vers la droite** : la queue s'allonge du côté des hauts salaires.
 3. Il n'y a **qu'une seule bosse** : on ne voit pas deux catégories distinctes de salariés.
- Ce troisième point est un diagnostic important. Deux bosses signaleraient deux populations mélangées — cadres et non-cadres, par exemple — qu'il faudrait étudier séparément.

Le polygone des fréquences

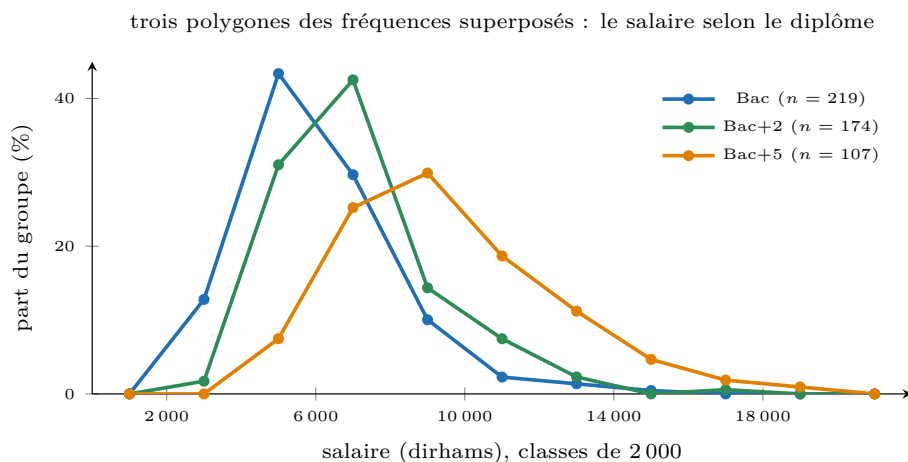
Construction

On joint par des segments les **milieux des sommets** des rectangles, et l'on ferme la ligne sur l'axe, au centre d'une classe vide de part et d'autre. L'aire sous le polygone reste égale à celle de l'histogramme.

Son usage : superposer

Deux histogrammes superposés se masquent l'un l'autre ; deux polygones se lisent ensemble. Le polygone sert donc à **comparer** deux distributions — deux entreprises, deux années, deux groupes. Pour des groupes d'effectifs différents, on trace les **fréquences** du groupe, pas les effectifs.

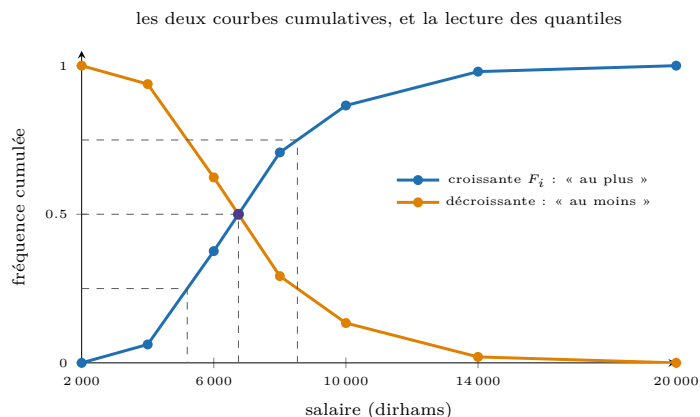
Trois groupes, trois polygones



Ce que la superposition révèle

Le polygone des Bac+5 est **décalé vers la droite** et **plus aplati** : salaires plus élevés et plus dispersés. Les groupes se chevauchent pourtant largement.

Les courbes cumulatives



les deux courbes se croisent à la **médiane** : autant de salariés au-dessus qu'en dessous.

Tout quantile se lit de la même façon : on entre par l'axe des fréquences, on sort par l'axe des salaires.

Interpolation linéaire dans la classe qui contient le seuil.

Comment les lire

Construction

- La courbe **croissante** joint les points $(e_i; F_i)$: elle part de 0 et monte vers 1.
- La courbe **décroissante** joint les points $(e_{i-1}; 1 - F_{i-1})$: elle part de 1 et descend vers 0.

Entre deux bornes, on relie par un **segment de droite** : c'est l'hypothèse d'équirépartition du chapitre 3.

Les deux sens de lecture

- On entre par l'**axe des salaires** et l'on sort par les fréquences : « quelle part gagne moins de 6 000 dirhams ? » Réponse : 37,6 %.
- On entre par l'**axe des fréquences** et l'on sort par les salaires : « au-dessous de quel salaire se situe la moitié du personnel ? » C'est la **médiane**, objet du chapitre 6.

Lire entre deux bornes

Quelle part gagne moins de 9 000 dirhams ?

9 000 n'est pas une borne : il se trouve au milieu de $[8\,000; 10\,000[$, où F passe de 0,708 à 0,866. Sur le segment qui joint ces deux points :

$$F(9\,000) = 0,708 + (0,866 - 0,708) \times \frac{9\,000 - 8\,000}{2\,000} = 0,708 + 0,079 = 0,787$$

Soit 78,7 %. Sur les données individuelles : 80,2 %.

La même règle de trois, dans les deux sens

Lire une fréquence pour un salaire donné, ou un salaire pour une fréquence donnée : c'est la **même interpolation linéaire**, faite sur le même segment, dans un sens ou dans l'autre.

Lire la médiane sur la courbe

La lecture de la médiane, par interpolation

La fréquence 0,5 est atteinte dans la classe [6 000 ; 8 000[, où F passe de 0,376 à 0,708.

$$Me = 6\,000 + 2\,000 \times \frac{0,500 - 0,376}{0,708 - 0,376} = 6\,000 + 2\,000 \times 0,373 = 6\,747$$

Valeur exacte sur les données individuelles : 6 770 dirhams. L'écart tient à l'hypothèse d'équirépartition.

Le point où les deux courbes se croisent

Elles se coupent là où « autant en dessous » égale « autant au-dessus », c'est-à-dire à la médiane. C'est le moyen le plus rapide de la repérer à l'œil sur un graphique publié.

Bien faire un graphique

Les règles

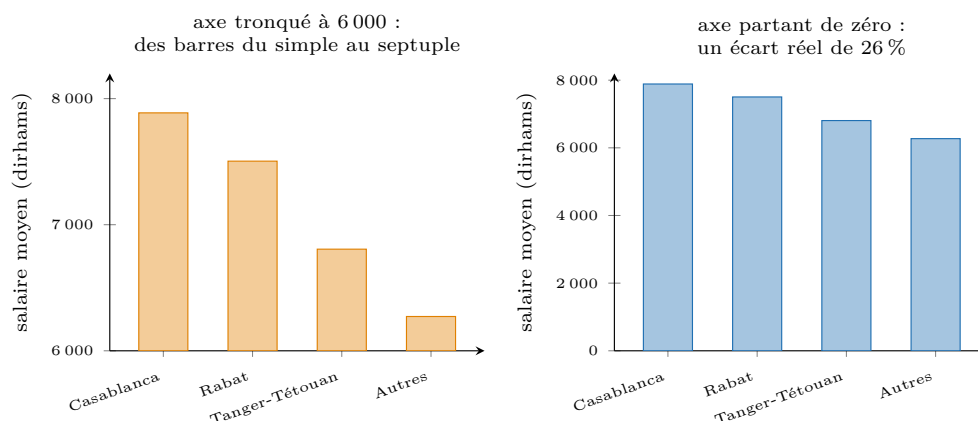
Cinq exigences

1. Un **titre** qui dit ce que le graphique montre.
2. Des **axes légendés**, avec les unités.
3. L'axe des ordonnées **part de zéro** — sauf mention explicite du contraire.
4. La **source** et l'**effectif** n .
5. Aucune décoration qui n'apporte pas d'information.

La règle 3 est la plus violée, et la plus grave

Tronquer l'axe des ordonnées multiplie visuellement l'ampleur d'une variation. C'est le procédé le plus courant des graphiques trompeurs — dans la presse, dans les rapports d'activité, et dans les mémoires d'étudiants. **Si vous devez tronquer l'axe, dites-le sur le graphique.**

L'axe tronqué, en image



Le calcul de la tromperie

Salairer moyen : 7 887 dirhams à Casablanca, 6 272 pour « Autres ». Rapport réel : $7\,887/6\,272 = 1,26$. Avec l'axe à 6 000, les barres mesurent 1 887 et 272 : rapport apparent 6,9.

Ni ombre, ni relief

Le principe général

Un graphique ne doit rien contenir qui ne serve à lire les données : ni effet d'ombre, ni troisième dimension, ni dégradé.

Un « camembert en 3D » déforme les aires par la perspective : les secteurs de devant paraissent plus grands. C'est un graphique **faux**, pas seulement laid.

La liste de contrôle avant de publier

Titre ? Unités ? Axe à zéro ? Source et n ? Type de graphique adapté au type de variable ? Pour un histogramme, hauteurs en densité si les classes sont inégales ? Une seule réponse « non », et le graphique est à refaire.

S'entraîner

Exercice 1 : le budget d'un ménage

Représenter des parts d'un tout

Un ménage de Tétouan consacre chaque mois 9 000 dirhams à ses dépenses :

Poste	Alimentation	Logement	Transport	Éducation	Autres
Dirhams	3 600	2 400	1 200	900	900

1. Calculer la fréquence de chaque poste et l'angle de son secteur.
2. Pourquoi le secteur convient-il ici ?
3. On veut comparer ce budget à celui d'un ménage de Casablanca. Quel graphique choisir, et pourquoi pas deux disques côte à côte ?

Exercice 1 : solution

Fréquences et angles

Poste	Alimentation	Logement	Transport	Éducation	Autres
f_i	0,400	0,267	0,133	0,100	0,100
Angle	144°	96°	48°	36°	36°

Contrôles : $\sum f_i = 1$ et $\sum \alpha_i = 360^\circ$.

Le choix du graphique

2. Il s'agit bien de **parts d'un tout**, en petit nombre (5) et nettement différentes.
3. Des **barres groupées**, poste par poste, en fréquences : on compare des longueurs alignées. Sur deux disques, il faudrait comparer des angles placés à des endroits différents — l'œil s'y trompe.

Exercice 2 : lire une courbe dans les deux sens

Les montants de crédits accordés par une agence

Montant (milliers de dh)	F_i en fin de classe
$[0; 50[$	0,20
$[50; 100[$	0,55
$[100; 200[$	0,85
$[200; 400[$	1,00

En reliant les points par des segments :

1. Quelle part des crédits est inférieure à 75 000 dirhams ?
2. Quelle part dépasse 150 000 dirhams ?
3. Au-dessous de quel montant se trouvent 80 % des crédits ?
4. Déterminer le montant médian.

Exercice 2 : solution

De la valeur vers la fréquence

1. $F(75) = 0,20 + (0,55 - 0,20) \times \frac{75 - 50}{50} = 0,20 + 0,175 = 0,375$, soit 37,5 %.
2. $F(150) = 0,55 + 0,30 \times \frac{50}{100} = 0,70$, donc $1 - 0,70 = 30\%$ des crédits dépassent 150 000 dirhams.

De la fréquence vers la valeur

3. 0,80 est atteint dans $[100; 200[$: $100 + 100 \times \frac{0,80 - 0,55}{0,85 - 0,55} = 100 + 83,3 = 183,3$ milliers de dirhams.
4. 0,50 est atteint dans $[50; 100[$: $Me = 50 + 50 \times \frac{0,50 - 0,20}{0,35} = 50 + 42,9 = 92,9$ milliers de dirhams.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. Chaque colonne du tableau appelle son graphique ; le choix dépend du **type de variable**.
2. **Bâtons séparés** pour le qualitatif et le discret, **rectangles jointifs** pour le continu.
3. **Secteurs** pour les parts d'un tout, en petit nombre : $\alpha_i = 360^\circ \times f_i$.
4. **Histogramme** : hauteur = densité, **aire** = effectif.
5. **Polygones des fréquences** : pour superposer et comparer des distributions.
6. **Courbes cumulatives** : deux sens de lecture, par interpolation ; elles se croisent à la médiane.
7. Axe des ordonnées à partir de **zéro**, sauf mention explicite.

La suite

Le tableau et les graphiques donnent la **forme**, pas des nombres synthétiques : impossible de comparer deux entreprises en superposant deux histogrammes. La partie III remplace la forme par quelques nombres, en commençant par la tendance centrale.