

Statistique descriptive

Chapitre 5 : Une colonne, un graphique



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Le principe

Les quatre graphiques

Bien faire un graphique

Ce qu'il faut retenir

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

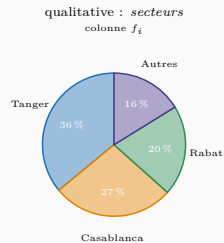
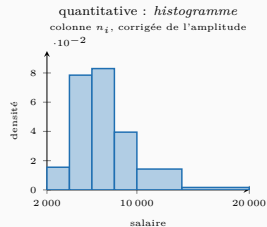
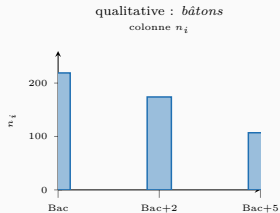
Chaque colonne appelle son graphique (suite)

Un graphique n'illustre pas : il montre

Un tableau donne les valeurs exactes; un graphique donne la **forme**. Ils ne répondent pas à la même question.

Le tableau répond à « combien ? ». Le graphique répond à « à quoi cela ressemble-t-il ? » — l'asymétrie, la présence de deux bosses, une valeur aberrante. Ces choses-là ne se voient pas dans une colonne de chiffres.

Les quatre graphiques



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 exactement 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. Un graphique à barres collées pour une variable qualitative est une faute de lecture, même s'il est joli.

Le tri des modalités

- Variable **ordinaire** : on respecte l'ordre naturel — Bac, Bac+2, Bac+5. Jamais autre chose.
- Variable **nominaire** : 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.

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°

Ses limites, qui sont 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.

L'œil humain compare bien des longueurs, mal des angles et très mal des aires. Pour comparer des proportions avec précision, le diagramme en barres est supérieur; le secteur ne garde qu'un avantage : il montre qu'on a bien affaire à un tout.

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 4).

Ce qu'on lit sur l'histogramme des salaires

Trois choses, immédiatement :

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** : la population est homogène, il n'y a pas deux catégories distinctes de salariés.

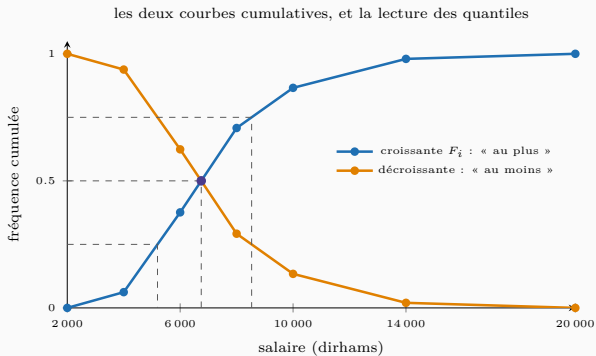
Ce troisième point est un diagnostic important. Une distribution à deux bosses signalerait deux populations mélangées — cadres et non-cadres, par exemple — qu'il faudrait étudier séparément.

Le polygone des fréquences

En joignant les **milieux des sommets** des rectangles, on obtient une ligne brisée : le **polygone des fréquences**.

Il sert surtout à **superposer** deux distributions sur un même graphique — comparer deux entreprises, deux années — ce que deux histogrammes superposés ne permettent pas.

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.

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 4.

Les deux sens de lecture

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

Le second sens est celui des quantiles, et c'est l'objet du chapitre 7.

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

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. Une hausse de 2 % peut être rendue spectaculaire en démarrant l'axe juste sous le premier point.

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.

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.

Ce qu'il faut retenir

Six 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 : la séparation encode le type.
3. **Secteurs** pour les parts d'un tout, et seulement si les modalités sont peu nombreuses.
4. **Histogramme** : hauteur = densité, **aire** = effectif.
5. **Courbes cumulatives** : deux sens de lecture, et elles se croisent à la médiane.
6. Axe des ordonnées à partir de **zéro**, sauf mention explicite.

La suite

Le tableau et les graphiques donnent la **forme**. Ils ne donnent pas de nombres synthétiques : impossible de comparer deux entreprises en superposant deux histogrammes.

La partie III remplace la forme par quelques nombres : c'est la statistique descriptive **univariée**, et elle commence par la tendance centrale.